OPEN SOURCE BIOINFORMATICS: THE INTERSECTION BETWEEN FORMAL INTELLECTUAL PROPERTY LAWS AND USER GENERATED LAWS IN THE SCIENTIFIC RESEARCH COMMONS

Thesis submitted to the Faculty of Law at the University of Tasmania for the degree of Doctor of Philosophy

by

James Scheibner, BComp-LLB (Hons), Grad Cert Research

Centre for Law and Genetics,University of Tasmania

November 2018

99,950 Words

Declaration of Originality

This thesis contains no material which has been accepted for a degree or diploma by the University of Tasmania or any other institution, except by way of background information and duly acknowledged in this thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of this thesis, nor does the thesis contain any material which infringes copyright

Authority of Access

This thesis may be made available for loan and limited copyright and communication in accordance with the Copyright Act 1968.

Statement of Ethical Conduct

The research associated with this thesis abides the international and Australian cods on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

DATE OF SUBMISSION 26 November 2018

SIGNED DATE - JAMES KEVIN SCHEIBNER

Open Source Bioinformatics: The Intersection between Formal Intellectual Property Laws and User Generated Laws in the Scientific Research Commons

This thesis examines the interplay between national copyright and patent laws, and informal user generated norms in the governance of open source bioinformatics projects. Bioinformatics describes an interdisciplinary merger between computer science, statistics and molecular biology for quantitative biology analysis. As for many computationally driven scientific fields, there is a strong initiative both by researchers inside and outside the field to encourage collaborative research through open source software and data licensing. However, a trend towards seeking exclusive copyright and patent protection for bioinformatics algorithms could foster exclusivity and discourage collaboration in bioinformatics.

Whether this effect exists within bioinformatics research, and if so is best resolved through national copyright and patent reform or private ordering strategies (which are already present through open source licensing) is a matter of open debate. This thesis explores these issues using a mixed methods, grounded theory framework that compares open licensing of bioinformatics software across the US, the EU, Australia and New Zealand. This framework operates on three levels: firstly, with a doctrinal analysis of copyright and patent laws (as well as related sui generis rights for data compilations); secondly, with a quantitative analysis of patent applications in bioinformatics and forward citation rates for patent publication pairs to determine whether the grant of these patents has a negative effect on citation rates; and thirdly, through semi-structured interviewing of bioinformaticians who release open source software and also seek patent protection.

Each of these layers of analysis reveals that national patent laws do not appear to have a significant effect on the formation and governance of open source bioinformatics communities, but law reform targeted at encouraging private ordering strategies through copyright licensing might have a positive effect. This thesis concludes by offering recommendations on assessing how private ordering strategies in open source bioinformatics can be improved to encourage collaborative research.

Acknowledgements

It feels as if a lifetime ago I decided to return to University and undertake a doctorate of laws. The process has been arduous and testing at times, but I need to offer some thanks to those who made it possible. First and foremost, I am indebted to my supervisors, Professor Dianne Nicol and Dr Jane Nielsen. From when I started my journey as a doctoral student to our last impromptu supervision meeting, they have always unconditionally supported me. In my darkest hours following an operation gone horribly wrong, they inspired me to keep going and finish what I had started. I credit my evolution as an academic researcher to their steady, patient guidance. I must also thank two research supervisors who were heavily involved at different stages of the PhD, namely Dr Janet Hope and Associate Professor Michael Charleston of the School of Physical Sciences. Janet was instrumental during the nascent stages of the thesis in helping me collect my ideas and prepare a cogent narrative. Likewise, Michael was instrumental in helping me develop an understanding of the field of bioinformatics.

I also must thank the academic and library staff at the Faculty of Law, who put up for me for far too long, first as an undergraduate student and later as a postgraduate student. In particular, my special thanks must go to Professor Don Chalmers, Professor Margaret Otlowski, Dr Lynden Griggs, Dr Elise Histed, Dr Jeremy Prichard, Associate Professor Heather Forrest and Dr Jeff McGee. I owe an enormous debt too to Professor Gino Dal Pont, who very generously read a draft of my thesis to ensure that the findings were coherent. I also must thank the incredibly tight knit and supportive postgraduate cohort at the Faculty of Law, who often patiently and kindly listened to the nervous mutterings of their anxious colleague. In particular, Dr Moshood Abdussalam, Jan Charbonneau, Dr John Liddicoat, Dr Kerryn Brent, Dr Philippa McCormack, Joseph Wenta and Dr Jason Allen Grant deserve special credit for their patience and sage advice. Last but most definitely not least, I must thank the tireless and dedicated library staff at the Law Library, particularly Chris Hurburgh, Carolyn Jarvis and Deb Bowering. Not only were they all very tolerant in handling my frequent inter-library loan requests, but they were also a constant source of good counsel and sympathy.

I must also thank the University of Tasmania Graduate Research Office, who granted me part of the McDougall Postgraduate Scholarship and the Sir Henry Baker Memorial Fellowship. These scholarships allowed me to attend the 2018 Conference of the Commons at the University of Utrecht, and the IP Academics Conference at the University of Sydney. These were invaluable experiences that allowed me to meet key researchers in my field and eventually complete my thesis. In addition, I must thank Linda Kahl, Tania Bubela and Osmat Jefferson, all of whom gave me excellent advice on the direction of my thesis. Although I cannot name them, I must also thank my interviewees who contributed to this project via their interview responses.

Finally, I must thank my friends and family who supported me during this thesis, particularly my brother Angus, Tim, Nick, Alex, India, Megan and Jez. Last, but certainly not least, I must thank my mother Tethys, who knows how hard completing a PhD is from personal experience, and worked tirelessly to help with drafting and editing my thesis.

The King said "The third question is, how many seconds of time are there in eternity." Then said the shepherd boy, "in Lower Pomerania is the Diamond Mountain, which is two miles and a half high, two miles and a half wide, and two miles and a half in depth; every hundred years a little bird comes and sharpens its beak on it, and when the whole mountain is worn away by this, then the first second of eternity will be over." - Household Tales, by the Brothers Grimm.

Contents

INTRODUCTION

The past decades have seen an increasing integration of computational techniques into research (otherwise known as data driven research[1]). Computational science has been instrumental in driving basic and applied biotechnology research, as we seek to improve our understanding of the natural world.[2] However, there is currently considerable debate amongst the legal and scientific community as to what intellectual property regimes are most appropriate for computational science.[3] In particular, there is a continuum of perspectives between strong proprietary approaches and more permissive commons based approaches that conform to traditional Mertonian norms of scientific research.[4] This debate is particularly pertinent for Australia. In 2015 the Federal and State governments spent 407 million Australian dollars (or 40 percent) of their total research and development budget on information communications technology research.[5]

This thesis seeks to examine the intersection between copyright and patent laws and informal user generated norms in the governance of open source bioinformatics projects. These copyright and patent laws include both national and international laws. Given that the purpose of this thesis is to fulfil the requirements of an Australian PhD, the primary focus is on Australia. However, recognising the cross-jurisdictional nature of computational research, the Australian analysis is undertaken within an international context. This introduction firstly briefly sets out the scientific, sociological, legal and economic background within which computational science has developed. This introduction then turns to an articulation of the

---

1 Aisling O'Driscoll, Jurate Daugelaite and Roy D. Sleator, "Big Data', Hadoop and Cloud Computing in Genomics' (2013) 46(5) *Journal of Biomedical Informatics* 774–781.

2 Harvey B. Newman, Mark H. Ellisman and John A. Orcutt, 'Data-Intensive E-Science Frontier Research' (2003) 46(11) *Communications of the ACM* 68–77 71-2.

3 Jerome H. Reichman, Tom Dedeurwaerdere and Paul F. Uhlir, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge University Press, 2016) 4.

4 Paul A. David, 'Understanding the Emergence of 'Open Science' Institutions: Functionalist Economics in Historical Context' (2004) 13(4) *Industrial and Corporate Change* 571–589 573.

5 Annette McLeod, *Returns on Investment: Considerations on Publicly Funded ICT Research and Impact Assessment* (PhD thesis, University of Melbourne, 2016) 21 <http://minerva-access.unimelb.edu.au/handle/11343/124272>.

research questions underpinning to this thesis. This introduction then concludes by describing methods and methodology adopted to answer these research questions and the scope and structure of this thesis.

1. Scientific Background

For the purpose of this thesis, the OECD definition of 'biotechnology' is adopted: the 'application of science and technology to living organisms as well as parts, products and models thereof, to alter living or non-living materials for the production of knowledge, goods and services'.[6] The term 'bioinformatics' is used to describe the merger of the academic fields of computer science and molecular biology, and describes software techniques used to support molecular biology research, with a specific emphasis on software used to analyse and compare genome sequences.[7] Bioinformatics software is an integral component of next generation sequencing (NGS) technology, which involves high throughput sequencing of genome and protein sequences.[8]

Bioinformatics represents a sub-discipline of computational biology, which refers to a broad approach of understanding biological systems using computer science techniques.[9] Under this broad definition there are a number of sub categories of computational biology. First, systems biology is closely related to bioinformatics but refers to modelling complex intercellular and intracellular processes using computational models and simulations. Systems biology builds on existing bioinformatics techniques by contextualising genomic expression in biological systems more broadly.[10] Synthetic biology is another sub-discipline of computational biology, and is concerned with the creation of artificial organisms or biological parts that function in predictable ways. Although each of these disciplines is a sub-discipline of computational biology, each was inspired by and continues to be influenced by advances in bioinformatics and computer science.[11] Jerome Reichman, Paul Ulhir and Tom

---

[6] *OECD Glossary of Statistical Terms - Biotechnology, Single Definition <https://stats.oecd.org/glossary/detail.asp?ID=219>*

[7] Paulien Hogeweg, 'The Roots of Bioinformatics in Theoretical Biology' (2011) 7(3) *PLOS Computational Biology* e1002021 2.

[8] Wilhelm J. Ansorge, 'Next-generation DNA sequencing techniques' (2009) 25(4) *New biotechnology* 195–203 165.

[9] Nicholas M. Luscombe, Dov Greenbaum and Mark Gerstein, 'What is Bioinformatics? A Proposed Definition and Overview of the Field' (2001) 40(4) *Methods of information in medicine* 346–358 356.

[10] Leroy Hood et al., 'Systems Biology and New Technologies Enable Predictive and Preventative Medicine' (2004) 306(5696) *Science* 640–643 641.

[11] Institute of Medicine (US) Forum on Microbial Threats, *The Science and Applications of Synthetic and Systems Biology: Workshop Summary* (National Academies Press (US), 2011) 2-3; Michael B. Elowitz and Stanislas Leibler, 'A Synthetic Oscillatory Network of Transcriptional Regulators' (2000) 403(6767) *Nature* 335–338;

Dedeurwaerdere consider these 'in silico' approaches as part of a broader trend in biology research. The nature of this trend is described more fully in Table 1.

Pre-1990s	**Recent Past and Future Trends**
Primary focus on single organisms and subsystems	Increasing focus on inter-dependent and complex systems
Intra-disciplinary	Interdisciplinary
Atomistic/insular/local	Integrative/collaborative/global
In vitro research	*In silico* research
Print communication	Networked digital communication
Data limited	'Big data', especially genomic data
'Small science' organisations that operate on a national level	Increasingly 'big science' organisations that operate on a supranational or international level
Strong distinction between basic research (largely publicly funded) and applied research (largely privately funded).	Blurring of the boundaries between publicly funded and privately funded research, as well as basic and use inspired research

Table 1: The Changing Characteristics of Molecular Biology Research. Adapted from a table in Jerome H Reichman, Tom Dedeurwaerdere and Paul F Uhlir, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons*, (Cambridge University Press, 2016), 13

This new approach to biology research coincided with the evolution of information systems theory and in particular the emergence of open source development in software engineering. 'Open source development' refers to a software engineering strategy that involves releasing the source code[12] along with the object code.[13] This release pattern allows others to modify and improve that software.[14] These development practices are frequently reinforced through open source licences, which mandate the requirements for attribution or relicensing of source code.[15] Open source software development and open source licensing both evolved from the free software movement. This movement was founded by MIT computer scientist Richard Stallman in response to professional software development companies seeking to commercialise software and curtain the nascent academic 'hacking' movement.[16] Stallman's free software movement had decidedly ideological goals (that is, liberating software from the clutches of proprietary protection). However, open source

---

Timothy S. Gardner, Charles R. Cantor and James J. Collins, 'Construction of a Genetic Toggle Switch in *Escherichia coli*' (2000) 403(6767) *Nature* 339–342.

[12] The underlying design for a software application.

[13] The executable component of the software

[14] Kevin Crowston et al., 'Free/Libre Open-Source Software Development: What We Know and What We Do Not Know' (2012) 44(2) *ACM Computing Surveys* 7–7:35 9.

[15] Josh Lerner and Jean Tirole, 'The Scope of Open Source Licensing' (2005) 21(1) *The Journal of Law, Economics, and Organization* 20–56 21.

[16] Johan Söderberg, *Hacking Capitalism: The Free and Open Source Software Movement* (Routledge, 2015) 19.

development and licensing were also driven by the practical need for reproducible research software in scientific research.[17] Johan Sonnenburg argues that the potential benefits of releasing scientific software under an open source licence can include the following:

1. reproducibility of scientific results and fair comparison of algorithms;
2. uncovering problems in source code;
3. building on existing software methods (rather than re-implementing them);
4. access to scientific methods without significant limitations;
5. combination of different software techniques;
6. faster adoption of methods in different disciplines and in industry; and,
7. collaborative emergence of standards.

2. Sociological Background

As Rebecca Eisenberg notes, 'functionalist accounts of scientific norms stress the cumulative and iterative nature of scientific discovery'.[18] As part of scientific progress, researchers can expand on or circumvent previous research. Conventional economic theory dictates that both copyright and patent law fulfil this function by protecting the appropriation of expressed or published works, correcting market failure. That is, copyright can only vest in expressed works as opposed to ideas or functional objects. Likewise, patents can only be granted for novel and useful inventions where the best means for performing that invention is disclosed within a patent application. In return, the successful grant of these rights allows a copyright or patent holder to recover the sunk costs associated with its development.[19]

However, emergent forms of technology, including computational biology, are challenging these conventional economic theories in a number of ways. Specifically, each of these fields handles an inherently intangible form of technology. Inventions that are both intangible and functional represent a challenge for both patent law, which has traditionally only protected

---

[17] Charles Schweik, Tom Evans and J. Morgan Grove, 'Open Source and Open Content: a Framework for Global Collaboration in Social-Ecological Research' (2005) 10(1) *Ecology and Society* 33.

[18] Rebecca S. Eisenberg, 'Patents and the Progress of Science: Exclusive Rights and Experimental Use' (1989) 56(3) *University of Chicago Law Review* 1017 1055.

[19] Thorsten Käseberg, *Intellectual Property, Antitrust and Cumulative Innovation in the EU and the US* (Bloomsbury Publishing, 2012) 49.

inventions that have 'some tie to physical materiality'[20] and copyright law, for which purely functional works have been traditionally excluded from the scope of protection.[21] Moreover, the inflexibility of traditional patent and copyright laws has been exacerbated by a trend that has been referred to as the 'dematerialisation' of these technologies. In other words the physical form of an invention[22] is increasingly separated from its informational content.[23] In concert, these two trends have created enormous difficulties for legislators and judges in determining the appropriate scope and overlap of copyright and patent law.[24]

These issues are complicated by the previously mentioned trend towards open source licensing. Open source advocates argue that either broadly defined copyright or patent rights have the potential to discourage developers from using or releasing open source licensed works.[25] However, there is equivocal evidence as to the impact of patents and copyright on innovation and open licensing more broadly. In particular, empirical studies suggest that the impact of patents and copyright on scientific research depends on the technological context but also the relevant legal regime and institutional norms.[26] Victoria Stodden argues that patent and copyright protection can successfully co-exist alongside open source licensing regimes.[27] This thesis attempts to assess whether patent and copyright protection have a positive, negative or neutral impact on the formation of open source communities in

---

[20] Jessica C. Lai, 'The Nebulous "Invention": From "Idea and Embodiment" to "Idea/Embodiment and Observable Physical Effects?"' in Antoinette Maget Dominicé and Jessica C. Lai (eds.), *Intellectual Property and Access to Im/material Goods* (Edward Elgar Publishing 2016)121 121.

[21] Dennis S. Karjala, 'Protecting Innovation in Computer Software, Biotechnology, and Nanotechnology' (2011) 16(1) *Virginia Journal of Law & Technology* 42–65 45.

[22] For example, the process for reading a genome sequence.

[23] For example, the information contained within genotypes that is responsible for the expression of particular phenotypes. (Charles Lawson and Michelle Rourke, 'Open Access DNA, RNA and Amino Acid Sequences: The Consequences and Solutions for the International Regulation of Access and Benefit Sharing' (2016) 24(3) *Journal of Law and Medicine* 96–118 99)

[24] In the United States (US) key cases on copyright include *Oracle America Inc v. Google Inc* 750 F.3d 1339 (Fed. Cir. 2014). In the European Union (EU) key cases on copyright include *SAS Institute Inc. v. World Programming Ltd* [2010] EWHC 1829 (Ch.) and Case C-393/09 *Bezpečnostní softwarová asociace - Svaz softwarové ochrany (BSA) v Ministry of Culture of the Czech Republic* [2010] ECR I-13971. In the US key cases on software patent law *Alice Corp. v. CLS Bank International* 134 S.Ct. 2347 (2014), *Mayo Collaborative Services v. Prometheus Laboratories Inc* 132 S. Ct. 1289 (2012), and *Association for Molecular Pathology v. Myriad Genetics Inc* 133 S. Ct. 2107 (2013). In the EU key cases on software patent law include T-208/84 (15th August 1986) [1987] O.J. EPO ('*Vicom/Computer-related invention*'); T-38/86 (14th February 1989) [1990] O.J. EPO ('*IBM/Text Processing*'); T-1173/97 (1st July 1998) [1997] O.J. EPO ('*Computer program product/IBM*'); T-935/97 (4th February 1999) EPOR ('*Computer program product II/IBM*') and *Symbian Ltd v Comptroller General of Patents* [2009] RPC 1.

[25] Marcus M. Dapp, *The Effects of Software Patent Policy on the Motivation and Innovation of Free and Open Source Software Developers* (PhD Thesis, ETH Zurich, 2009) 53-4; Clark D. Asay, 'Software's Copyright Anticommons' (2016) 66(2) *Emory Law Journal* 265–332 268.

[26] Zhen Lei, Rakhi Juneja and Brian D. Wright, 'Patents versus Patenting: Implications of Intellectual Property Protection for Biological Research' (2009) 27(1) *Nature Biotechnology* 36–40 38.

[27] Victoria Stodden, 'What Computational Scientists Need to Know about Intellectual Property Law' in *Implementing Reproducible Research* (Chapman and Hall/CRC 2014) 330-4.

computational biology research.

3. Legal and Economic Background

Due to the well-worn pathway between basic and applied molecular biology and pharmaceutical research more broadly, researchers and research sponsors have become eager to translate bioinformatics research into applied settings.[28] These settings including using bioinformatics in personalised and precision medicine, as well as pharmacogenomics.[29] In addition, bioinformatics and related information technology may play a critical role in the development of commercial biofuels or the production of drought resistant crops.[30] One way in which biomedical or biotechnology inventions are protected and commercialised is through strategic patent acquisition. In theory, these patents encourage private investment and exclusive cross licensing to establish research collaborations.[31]

There are concerns that patents (and to a lesser extent copyright)[32] may either intentionally or inadvertently be used to prevent competitors from engaging in research. These effects may eventuate through 'rent seeking' (charging exorbitant fees for use) or through the application of onerous terms of use.[33] Michael Heller argues in his 'tragedy of the anti-commons' hypothesis that fragmented property rights have the potential to prevent the use of a single resource to solve a collective action problem. This fragmentation therefore frustrates an otherwise socially desirable outcome.[34] Heller, in concert with Rebecca Eisenberg, explains that biotechnology research is particularly vulnerable to this effect due to the highly cumulative nature of applied biotechnology research. This effect increases the difficulty in

---

[28] Rebecca S. Eisenberg, 'Proprietary Rights and the Norms of Science in Biotechnology Research' (1987) 97(2) *Yale Law Journal* 177–231 204-5; Christopher M. Holman, Claes Gustafsson and Andrew W. Torrance, 'Are Engineered Genetic Sequences Copyrightable?: The U.S. Copyright Office Addresses a Matter of First Impression' (2016) 35(3) *Biotechnology Law Report* 103–111 106.

[29] J. Andreu-Perez et al., 'Big Data for Health' (2015) 19(4) *IEEE Journal of Biomedical and Health Informatics* 1193–1208 1196-7.

[30] Robert T. Furbank and Mark Tester, 'Phenomics – Technologies to Relieve the Phenotyping Bottleneck' (2011) 16(12) *Trends in Plant Science* 635–644 635, 637; Michael Halewood et al., 'Plant Genetic Resources for Food and Agriculture: Opportunities and Challenges Emerging from the Science and Information Technology Revolution' (2018) 217(4) *New Phytologist* 1407–1419 411.

[31] Jeanne C. Fromer, 'Patent Disclosure' (2008) 94(2) *Iowa Law Review* 539–606 559.

[32] Robin Feldman and John Newman, 'Copyright at the Bedside: Should We Stop the Spread?' (2012) 16(3) *Stanford Technology Law Review* 623–656 652.

[33] Tania Bubela, Jenilee Guebert and Amrita Mishra, 'Use and Misuse of Material Transfer Agreements: Lessons in Proportionality from Research, Repositories, and Litigation' (2015) 13(2) *PLoS Biol* e1002060 6-7.

[34] Michael A. Heller, 'The Tragedy of the Anticommons: Property in the Transition from Marx to Markets' (1997) 111(3) *Harvard Law Review* 621–688 624.

co-ordinating biotechnology research patents due to the high transaction costs of bundling multiple licences.[35]

Despite Heller and Eisenberg's dire predictions, evidence from within the biotechnology research sector (as well as in academic research) is contextual. This context specificity holds particularly true for areas of research where patents have minimal importance, such as computer science.[36] In particular, there have been strong community led initiatives to encourage open source software development practices within the broader biotechnology research community. As discussed in Section 1, the emergence of computational biology represents a shift in molecular biology research towards a 'data oriented' approach.[37] This shift towards computational research has also been reflected in increased scientific collaboration through networked technology. Finally, much of the earlier innovation with respect to bioinformatics algorithms has occurred in the absence of either patent acquisition or the use of exclusive proprietary licensing.[38]

Nevertheless, there is increasing pressure to apply bioinformatics and computational biology research into commercially viable research.[39] The purpose of this research translation is ostensibly to harness the enormous potential of computational science and direct it towards applied scientific research. This pressure has been coupled with the establishment of dedicated bioinformatics classes under both the International Patent Classification (IPC) and the United States Patent Classification (USPC) regimes.[40] Accordingly, bioinformatics represents a field where proprietary models are being overlaid onto a research field which has historically depended on open research norms.[41] This approach can be contrasted with other 'open source'

---

[35] Michael A. Heller and Rebecca S. Eisenberg, 'Can Patents Deter Innovation? The Anticommons in Biomedical Research' (1998) 280(5364) *Science* 698–701 700.

[36] John P. Walsh, Charlene Cho and Wesley M. Cohen, 'View from the Bench: Patents and Material Transfers' (2005) 309(5743) *Science* 2002–2003 2003-4; Michael Noel and Mark Schankerman, 'Strategic Patenting and Software Innovation' (2013) 61(3) *The Journal of Industrial Economics* 481–520 485.

[37] Jason E. Stajich and Hilmar Lapp, 'Open Source Tools and Toolkits for Bioinformatics: Significance, and Where Are We?' (2006) 7(3) *Briefings in Bioinformatics* 287–296 287-9.

[38] Mark Harvey and Andrew McMeekin, *Public Or Private Economies of Knowledge?: Turbulence in the Biological Sciences* (Edward Elgar, 2007) 71-8; Mark Harvey and Andrew McMeekin, 'Public or Private Economies of Knowledge: The Economics of Diffusion and Appropriation of Bioinformatics Tools' (2009) 4(1) *International Journal of the Commons* 481–506 485.

[39] Bruce Rasmussen, *Creating and capturing value in the biopharmaceutical sector* (PhD Thesis, Victoria University, 2008) 197 <http://www.vu.edu.au/research>.

[40] Hyun-Seok Park, 'Preliminary Study of Bioinformatics Patents and Their Classifications Registered in the KIPRIS Database' (2012) 10(4) *Genomics & Informatics* 271–274 273; Saurabh Vishnubhakat and Arti Rai, 'When Biopharma Meets Software: Bioinformatics at the Patent Office' (2015) 29(1) *Harvard Journal of Law & Technology* 206 224-5.

[41] Dov Greenbaum, 'Patent Sharing in Biotechnology' in Jorge L. Contreras and Meredith Jacob (eds.), *Patent Pledges: Global Perspectives on Patent Law's Private Ordering Frontier* (Edward Elgar Publishing 2017)56–81 56.

movements in biotechnology research, where researchers are attempting to introduce open innovation strategies into a research field where proprietary models of innovation dominate.[42]

4. Research Questions

This thesis is centred around answering the following research question: what aspects of patent and copyright protection, legal or otherwise, may positively or negatively influence the governance of open source bioinformatics projects? From this research question, three sub-research questions flow:

1. What is the extent of copyright and patent protection with respect to bioinformatics and computational biology software?
2. Are bioinformatics researchers relying on patent and copyright protection for open source bioinformatics and computational biology development?
3. Do bioinformatics researchers have positive or negative perspectives on copyright and patent protection for bioinformatics software?

The first sub question is addressed through a comparative doctrinal analysis of divergent copyright and patent laws in the US, the EU and New Zealand, in addition to Australia. The US and EU were chosen for examination in this thesis because these two jurisdictions represent the two largest research centres for software and biotechnology research. In addition, both of these jurisdictions have played host to the majority of the case law on the scope of patent and copyright subject matter.[43] New Zealand was chosen because of its comparative size with and proximity to Australia. Both are small and medium sized economics which nevertheless engage in a disproportionately high level of innovative activity.[44] Accordingly, the benefits of studying Australia and New Zealand are two fold. Firstly, a doctrinal analysis will inform developers about the scope of their ability to collaborative effectively both domestically and internationally. Secondly, it may also reveal how the Australian and New Zealand researchers might best structure their research communities through institutional norms and funding models to support open licensing. However, jurisdictional differences cannot explain every difference in technology transfer and patenting policy for computational biology software. This thesis explores other differences with the aid of a mixed methods methodology. The next section of this introduction will address the methodology and methods used to assess this effect.

---

[42] Christine Årdal and John-Arne Røttingen, 'Open Source Drug Discovery in Practice: A Case Study' (2012) 6(9) *PLoS Neglected Tropical Diseases* e1827 2.

[43] Geertrui Van Overwalle, 'Policy Levers Tailoring Patent Law to Biotechnology: Comparing U.S. and European Approaches Bend or Break - The Patent System in Crisis' (2011) 1(2) *UC Irvine Law Review* 433–514 439.

[44] Susy Frankel, *Test Tubes for Global Intellectual Property Issues* (Cambridge University Press, 2015) 14-18.

5. Methods and Methodology

To answer each of the research questions described above, this thesis uses a grounded theoretical framework along with a mixed methods approach. Grounded theory is a systematic methodology developed by Barney Glaser and Anselm Strauss for socio-economic research that involves synthesising novel theory through gathering data.[45] Other socio-economic research methodologies involve selecting a research framework, gathering data, and excluding data that does not pertain to that research framework. However, grounded theory involves a continuous process of gathering data to construct theory about an emergent sociological phenomenon. Within the grounded theory framework, there is a further divergence into two sub-schools of thought. On the one hand, Glaser's approach to grounded theory analysis, mandates that theory should only be reviewed post data analysis. On the other hand, Strauss's approach to data analysis, which was later co-authored with Juliet Corbin, suggests researchers engage with other theory to help guide grounded theory analysis.[46]

For this reason, Strauss and Corbin's grounded theory approach was adopted for this thesis, and was combined with the Knowledge Commons Framework.[47] Brett Frischmann, Michael Madison and Katherine Strandburg developed the Knowledge Commons Framework as an adaptation of Elinor Ostrom's Institutional Analysis and Design (IAD) framework.[48] The IAD framework was originally developed as a means for understanding the governance of collective action regimes for the use of natural resources (such as shared irrigation systems in Nepal or lobster fisheries in Maine).[49] Specifically, Ostrom argued that successful commons style arrangements featured the following seven types of rules:[50]

1. *Position rules* articulate what roles people play in the project;
2. *Boundary rules* define who is eligible for a position, the process of how a person is assigned to that position and rules related to how the person leaves that position;

---

[45] Barney G. Glaser and Anselm L. Strauss, *Discovery of Grounded Theory: Strategies for Qualitative Research* (Routledge, 2017) 168.

[46] Juliet M. Corbin and Anselm Strauss, 'Grounded Theory Research: Procedures, Canons, and Evaluative Criteria' (1990) 13(1) *Qualitative Sociology* 3–21 15.

[47] Amrita Mishra and Tania Bubela, 'Legal Agreements and the Governance of Research Commons: Lessons from Materials Sharing in Mouse Genomics' (2014) 18(4) *OMICS: A Journal of Integrative Biology* 254–273 260.

[48] Michael J. Madison, Brett M. Frischmann and Katherine J. Strandburg, 'Constructing Commons in the Cultural Environment' (2009) 95(4) *Cornell Law Review* 657–710 664-5.

[49] Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, 1990) 182-216.

[50] Elinor Ostrom and Sue E. S. Crawford, 'Classifying Rules' in Elinor Ostrom (ed.), *Understanding Institutional Diversity* (Princeton University Press 2009)168 186-208.

3. *Choice rules* define actions that project members can and cannot make;

4. *Aggregation rules* articulate the process for how conflict should be resolved;

5. *Information rules* specify how and what kind of information flows between group members and other interested parties;

6. *Payoff rules* assign some kind of reward or sanction to specific actions or outcomes.

7. *Scope rules* specify which outcomes may, must or must not be affected or produced in a given situation.

By contrast, the Knowledge Commons Framework was developed for understanding the collective governance of 'knowledge' resources. These resources include research data, open source software and online encyclopaedias. Compared to a natural commons, there is a divergence in both what amounts to a failure of the commons and applicable property regimes (namely intellectual property rights as opposed to real property rights). Open source bioinformatics projects share elements of two 'established' knowledge commons phenomena (namely open source software and scientific data repositories). The Knowledge Commons framework represents an appropriate theoretical framework to answer the questions posed in this thesis.[51]

A mixed methods approach was adopted because of the ability to triangulate different methods and overcome the weaknesses of solely relying on quantitative or qualitative methods.[52] As discussed by Shubha Ghosh, copyright and patent statutes and case law do not provide comprehensive guidance as to protection of software invention. In addition, there may be other factors which dictate whether these rights have a positive, negative or neutral effect on collaborative research.[53] Further, small scale qualitative sampling can provide rich data about a particular case study, or can be useful for establishing causal links for theory building.[54] However, qualitative results can be difficult to use to posit or evaluate general relationships about particular socio-economic phenomena. Likewise, quantitative methods can

---

[51] Siobhán O'Mahony, 'Guarding the Commons: How Community Managed Software Projects Protect their Work' (2003) 32(7) *Research Policy* 1179–1198 1182-3; Markus Perkmann and Henri Schildt, 'Open Data Partnerships between Firms and Universities: The Role of Boundary Organizations' (2015) 44(5) *Research Policy* 1133–1143 1136.

[52] Evan S. Lieberman, 'Nested Analysis as a Mixed-Method Strategy for Comparative Research' (2005) 99(3) *American Political Science Review* 435–452 435.

[53] Shubha Ghosh, 'How to Build a Commons: Is Intellectual Property Constrictive, Facilitating, or Irrelevant?' in *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press 2007)209–245 209-210.

[54] Barbara Hanson, 'Wither Qualitative/Quantitative?: Grounds for Methodological Convergence' (2008) 42(1) *Quality & Quantity* 97–111 100-1.

be used to demonstrate the plausibility of particular theories, but may not be useful in determining the causal relationships underlying those theories.[55]

To achieve methodological triangulation, this thesis relies on a combination of comparative copyright and patent doctrinal research, quantitative patent landscaping and qualitative semi-structured interviewing. The benefits of this approach are as follows. The doctrinal analysis phase involves combining a historical review of bioinformatics as a field with charting the boundaries of copyright and patent protection for software on a comparative jurisdictional basis. This analysis helps determine what parts of bioinformatics research are either copyrightable or patentable. In addition, this doctrinal analysis allows a consideration of the different ways that open source licences have been interpreted by US, European Union (EU), Australian and New Zealand courts. Crucially, there is limited case law on open source licence enforcement in both Australia and New Zealand. Therefore, this thesis examines these licences within a global context by drawing on US and EU legislation and case law to assist with the interpretation of Australian and New Zealand law.

Patent landscaping is then introduced as a quantitative research method to assess what bioinformatics patents were being filed by research institutes. This patent landscape is restricted to the United States Patent and Trademark Office (USPTO), the European Patent Office (EPO), the Australian Patent Office (APO) and the Intellectual Property Office of New Zealand.[56] This patent landscaping model is further restricted to research institutes to assess the conflict between open norms traditionally associated with scientific research and research commercialisation strategies at academic institutes.[57] Finally, from the series of patents that had been identified, a set of patent publication pairs are extracted. The forward citation rates for these articles are compared to determine whether there was a decline in citations following the grant of a patent.

However, this quantitative analysis cannot alone inform whether patents or proprietary licensing have an anti-commons effect within bioinformatics research. Accordingly, semi-structured interviews were used to assess how bioinformatics developers and researchers rely on copyright and patent law to govern open source bioinformatics projects. These interviewees were sampled on the basis of their status as patent applicants or via their

---

[55] Amy R. Poteete, Marco Janssen and Elinor Ostrom, *Working Together: Collective Action, the Commons, and Multiple Methods in Practice* (Princeton University Press, 2010) 12.

[56] Tania Bubela et al., 'Patent Landscaping for Life Sciences Innovation: Toward Consistent and Transparent Practices' (2013) 31(3) *Nature Biotechnology* 202–206 204.

[57] Partha Dasgupta and Paul A. David, 'Toward a New Economics of Science' (1994) 23(5) *Research Policy* 487–521 493; Victoria Stodden, 'The Scientific Method in Practice: Reproducibility in The Computational Sciences' (MIT Sloan School Working Paper No 4773-10, MIT Sloan School of Management, 2010) 20-1 <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1550193>.

publication records within bioinformatics research.[58] These interviews were then coded deductively and inductively using the Institutional Grammar Tool (IGT) metric. Ostrom and Sue Crawford developed the IGT to classify institutional arrangements used for the governance of commons-based resources depending on whether they are rules, norms or shared strategies.[59] In addition, these rules were classified depending on what category of IAD rule type they related to. From these results, this thesis provides recommendations on private ordering, funding and law reform strategies that can be used to encourage open source development practices in bioinformatics.

6. Scope of Research and Original Contribution to Knowledge

There has been substantial doctrinal analysis of open source software licensing in the US and Europe.[60] However, there are comparatively fewer studies considering the enforceability of open source licences outside of the US and the EU, such as in Australia and New Zealand.[61] In particular, previous studies have largely focused on the legal risks associated with the use of open source software. However, they do not consider the enforceability of rules used to control exit and entry into open source communities. Likewise, there has been significant socio-legal research in the US into the use of collaborative licensing and patent pooling to create different research commons arrangements.[62] However, as identified by Geertrui van Overwalle, the question of whether jurisdictional differences in patent or copyright laws influence research commons success remains unexplored.[63]

---

[58] Dietmar Harhoff, Frederic M Scherer and Katrin Vopel, 'Citations, Family Size, Opposition and the Value of Patent Rights' (2003) 32(8) *Research Policy* 1343–1363 1348; Amrita Mishra, Paul N. Schofield and Tania M. Bubela, 'Sustaining Large-Scale Infrastructure to Promote Pre-Competitive Biomedical Research: Lessons from Mouse Genomics' (2016) 33(2) *New Biotechnology* 280–294 283.

[59] Sue E. S. Crawford and Elinor Ostrom, 'A Grammar of Institutions' (1995) 89(3) *American Political Science Review* 582–600 583-4, 588.

[60] Andres Guadamuz Gonzalez, 'Open Science: Open Source Licenses in Scientific Research' (2005) 7(2) *North Carolina Journal of Law & Technology* 321–366; Mikko Välimäki, 'Copyleft Licensing and EC Competition Law' (2006) 27(3) *European Competition Law Review* 130–136; Heather J. Meeker, 'Open Source and the Age of Enforcement' (2012) 4(2) *Hastings Science & Technology Law Journal* 267–290.

[61] Nevertheless, there are some limited examples in this case (Brian Fitzgerald and Nic Suzor, 'Legal Issues for the Use of Free and Open Source Software in Government' (2005) 29(2) *Melbourne University Law Review* 412; Susan Corbett, 'Creative Commons Licences, the Copyright Regime and the Online Community: Is there a Fatal Disconnect?' (2011) 74(4) *The Modern Law Review* 503–531).

[62] Jerome H. H. Reichman and Paul F. Uhlir, 'A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment' (2003) 66(1/2) *Law and Contemporary Problems* 315–462 429-30; Jerome H. Reichman and Ruth L. Okediji, 'When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale' (2012) 96(4) *Minnesota Law Review* 1362–1480.

[63] Geertrui Van Overwalle, 'Governing Genomic Data: Plea for an "Open Commons"' in Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press 2014)137–154 151.

This thesis attempts to combine these separate lines of inquiry by synthesising a grounded theory around open source bioinformatics research. Rather than focusing on individual projects as case studies, this thesis examines the jurisdictional effects of divergent copyright, patent and technology transfer laws on open source bioinformatics projects.[64] In particular, the explicit constant comparison that is inherent in grounded theory enables this comparison of divergent copyright and patent regimes.[65] Grounded theory explicitly allows for a consideration of the impact of dematerialisation on genomic materials and information on the relevance and utility of patent and copyright laws to bioinformatics technology.[66]

7. Thesis Outline

In developing a grounded theory of the interaction between open source development of bioinformatics software and copyright patent protection, this thesis adopts the following structure:

Chapter One is framed as a historical review, and reviews the relevant literature on the development of bioinformatics software and the evolution of computational biology as a research field. Chapter One discusses how molecular biology converged towards informational approaches whilst computer science became formalised as an academic discipline in the latter half of the twentieth century. This merger is framed in the context of large-scale computational biology projects such as the Human Genome Project and the International HapMap Project. Chapter One then discusses more recent developments in computational biology, such as the flood of data from Next Generation Sequencing (NGS) and the emergence of synthetic and systems biology as sub-disciplines. This proliferation of computational biology techniques is then framed in the context of trying to find a balance between encouraging the reproducibility and the sustainability of these technologies. This conflict then frames the next two chapters regarding the scope of patent and copyright laws.

Chapter Two is the first of the doctrinal analysis chapters, and considers the scope of copyright protection in the US, the EU, Australia and New Zealand. This chapter first considers copyright under international law, including the scope of copyright subject matter as well the extent of exclusive rights granted to copyright holders and exceptions to those rights. It then provides

---

[64] Nicola Lucchi, 'Understanding Genetic Information as a Commons: From Bioprospecting to Personalized Medicine' (2013) 7(2) *International Journal of the Commons* 313–338 317.

[65] Siobhán O'Mahony, above n51, 1183; Jason Michael Davis, *Reconsidering Antarctic Bioprospecting through Territorialities of Science, Property, and Governance* (PhD Thesis, The Ohio State University, 2011) 55 <https://etd.ohiolink.edu/pg_10?0::NO:10:P10_ACCESSION_NUM:osu1299535648>.

[66] Antoine Danchin, 'Information of the Chassis and Information of the Program in Synthetic Cells' (2009) 3(1-4) *Systems and Synthetic Biology* 125–134 130.

a historical account of how copyright protection became the foremost means of protecting software. It then concludes by charting how the scope of software copyright protection has been legislated and interpreted in the US, the EU, Australia and New Zealand.

Chapter Three is the second doctrinal analysis chapter, and considers the scope of patent protection for software and other intangible subject matter in the US, the EU, Australia and New Zealand. This chapter starts with a discussion of the evolution of patent law, and explains how intangible subject matter became patentable. In particular, this chapter compares the historically broad standards of patentable subject matter in the US with the narrow scope of patentable subject matter in the EU, Australia and New Zealand, as well as the lack of specific patent exemptions in the US versus defined research exemptions in the EU, Australia and New Zealand.

Chapter Four considers what aspects of patent and copyright law extend to bioinformatics, including sequencing software, sequencing hardware and sequence data. This chapter then examines how copyright and patent law may not only provide incentives for conducting bioinformatics research but in concert with restrictive licensing may inhibit research. The second part of this chapter considers different theoretical frameworks for the analysis of open innovation in scientific research. It then explains why the Knowledge Commons framework is the most appropriate theoretical framework for this thesis. This chapter then concludes by examining the relationship between the Knowledge Commons framework and intellectual property rights in the context of open source bioinformatics projects.

Chapter Five is the first of the empirical chapters and describes the patenting landscape methodology that was used to assess patenting activity by bioinformatics researchers. Specifically, this chapter focuses on patents that were filed by and granted to research institutes. These patents are then analysed via their patent class to determine the main areas of patent activity. Out of these patents, a set of patent publication pairs (that is, publications where the disclosed technology has been subsequently patented) are identified. The citation rates for these articles are then analysed to determine whether the grant of a patent led to a decline in forward citations.

Chapter Six is the second of the empirical chapters and describes the semi-structured interview method that was used to interview researchers involved in bioinformatics in the US, the EU, Australia and New Zealand. This chapter first describes how these researchers were identified via their publication records or their status as a listed inventor on a patent application or patent grant. Chapter Six then describes how these interviews were conducted and how the interview transcripts were coded using the IGT. Finally, this chapter provides a preliminary discussion of the results flowing from these interviews. These results will influence the private ordering and law reform strategies described in Chapter Seven.

Chapter Seven evaluates the findings from the doctrinal and empirical components of this thesis. Specifically, this chapter is split into two parts. The first part of this chapter provides an evaluation of on how open source bioinformatics projects might be better structured to encourage open source development. The second part of this chapter assesses the need for law reform in Australia, and provides recommendations for a set of specific legislative reforms to accommodate open source development of bioinformatics project. The purpose of legislative reform is to ensure, to the greatest extent possible, that open source licences operate as consistently as possible across all jurisdictions. Without specific precedent on open source enforcement in Australia, the comparative jurisdictional approach undertaken in this thesis assists in defining the types of law reform that might better support open source developers. This comparative jurisdictional approach is supplemented by the cross-jurisdictional approach to the empirical research component of this thesis described in Chapters Five and Six.

The following conference presentations were delivered from material published in this thesis:

James Scheibner 'Why is Open Source licensing prevalent in Bioinformatics? A Study of the Relationship between Governance Frameworks and Intellectual Property Rights in the Development of Scientific Research Software', paper presented at the International Association for the Study of the Commons XVI 2017 Conference, 10-14 July 2017, Utrecht, Netherlands.

James Scheibner 'A Software Interoperability Exemption in Australian Patent Law: Compliance with Article 30 of the TRIPS Agreement', paper presented at the Australian IP Academics Conference, 20-21 July 2017, Sydney, Australia.

James Scheibner 'Reframing the Software and Business Method Patents Debate around Disclosure and Reuse', paper presented at the Association for Teachers and Researchers in Intellectual Property Conference, 23-26 October 2017, Wellington, New Zealand.

Chapter 1

A HISTORICAL AND TECHNICAL OVERVIEW OF COMPUTATIONAL BIOLOGY

1.1 INTRODUCTION

This chapter introduces the foundational technical concepts in software engineering, computational biology and bioinformatics via a historical review.[1] These technical concepts will frame the doctrinal analysis of intellectual property rights and their relationship to open approaches to computational biology in later chapters of this thesis. In addition, this chapter introduces the primary justifications for these open approaches to computational biology, particularly within pure or use inspired basic research.[2] In this context, *basic research* involves acquiring foundational knowledge through underlying phenomenon and observations. By contrast, *applied research* involves developing new knowledge and particular applications of knowledge.

Although this thesis poses legal questions, it sits at an interdisciplinary crossroads. Whilst this thesis specifically focuses on computational biology, it also draws on information systems and technology management literature to explain how open source communities form.[3] These studies have also considered how different stakeholders interact with each other in scientific software development.[4] The theory flowing from these case studies will be used to inform how rules for the sharing of software and data form in computational biology communities.[5] This emphasis on institutional arrangements turn relies on institutional economics theory, which attempts to discern how different actors interact to manage a shared resource.[6]

To this end, this chapter is split into four parts. Section 1.2 explores the evolution of experimental molecular biology as a discipline during the 20th century. Section 1.2 contrasts this experimental approach with the natural history-oriented approach that has dominated in the study of biological sciences. This section documents the rise of experimental molecular biology techniques, particularly statistical techniques, as methods for understanding biological systems. Section 1.3 then covers parallel developments in computer science and software engineering. In particular, this section documents a shift in software development practices away from monolithic, hardware specific development towards modular, hardware

---

[1] Diane Vaughan, 'Theorizing Disaster Analogy, historical ethnography, and the Challenger accident' (2004) 5(3) *Ethnography* 315–347 312.

[2] Jason E. Stajich and Hilmar Lapp, 'Open Source Tools and Toolkits for Bioinformatics: Significance, and Where Are We?' (2006) 7(3) *Briefings in Bioinformatics* 287–296 288.

[3] Martin Michlmayr, 'Community Management in Open Source Projects' (2009) X(3) *The European Journal for the Informatics Professional, X (3)* 22–26 22.

[4] James Howison and James D. Herbsleb, 'Scientific Software Production: Incentives and Collaboration' (Paper presented at *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* , 2011).

[5] Amrita Mishra, Paul N. Schofield and Tania M. Bubela, 'Sustaining Large-Scale Infrastructure to Promote Pre-Competitive Biomedical Research: Lessons from Mouse Genomics' (2016) 33(2) *New Biotechnology* 280–294.

[6] Nitin Aggarwal and Eric A. Walden, 'Intellectual Property Bundle (IPB) theory: Managing Transaction Costs in Technology Development Through Network Governance' (2009) 48(1) *Decision Support Systems* 23–32.

independent software development. This section concludes by examining the socio-technical factors that led to the spread of software as an important research tool in scientific research.

Section 1.4 explores the merger of these two fields in computational biology. In particular, it explains how the introduction of quantitative methods into experimental molecular biology cemented an informational approach to understanding organic molecules and genomics, which in turn encouraged molecular biologists to apply computational tools to analyse sequences, including sequence collections and comparison algorithms. In addition, this section highlights how improvements in networking technology drove the first large scale genomics initiatives, such as the Human Genome Project (HGP) and the International HapMap Project. This section concludes by explaining the diverse current applications of computational biology, focusing in particular on biomedical and agricultural research. Finally, Section 1.5 offers preliminary conclusions regarding the nature of computational biology, particularly the two competing considerations of reproducibility and sustainability. The doctrinal and empirical analysis in this thesis of open source computational biology models is framed around these concepts.

1.2 THE PRE-HISTORY OF BIOINFORMATICS - FROM 'OLD BIOLOGY' TO 'NEW BIOLOGY'

1.2.1 *Defining Computational Biology and Bioinformatics*

As a preliminary matter, it is important to address what computational biology and bioinformatics. The theoretical biologists Paulien Hogeweg and Ben Hesper are attributed with first defining computational biology as 'the study of informatic processes in biotic systems'.[7] This definition can be contrasted with genetics (the study of how traits and genes for traits are passed between different organisms) and molecular biology (the study of molecular level biology).[8] Figure 1.1, reproduced from Professor Christopher Burge provides an overview of the relationship between different sub-disciplines of computational biology.

This diagram demonstrates the distinction between bioinformatics tools and computational biology as a scientific discipline. David Searls characterises this distinction as the difference between 'the science that informs the tools' and the 'tools that enable the science'.[9] In arguing against the concept of bioinformatics as a separate discipline, Searls posits that the emergence

---

[7] Paulien Hogeweg, 'Simulating the Growth of Cellular Forms' (1978) 31(3) *Simulation* 90–96; Paulien Hogeweg and Ben Hesper, 'Interactive Instruction on Population Interactions' (1978) 8(4) *Computers in Biology and Medicine* 319–327.

[8] Paulien Hogeweg and Ben Hesper, above n7; Paulien Hogeweg, 'The Roots of Bioinformatics in Theoretical Biology' (2011) 7(3) *PLOS Computational Biology* e1002021.

[9] David B. Searls, 'The Roots of Bioinformatics' (2010) 6(6) *PLOS Computational Biology* e1000809.

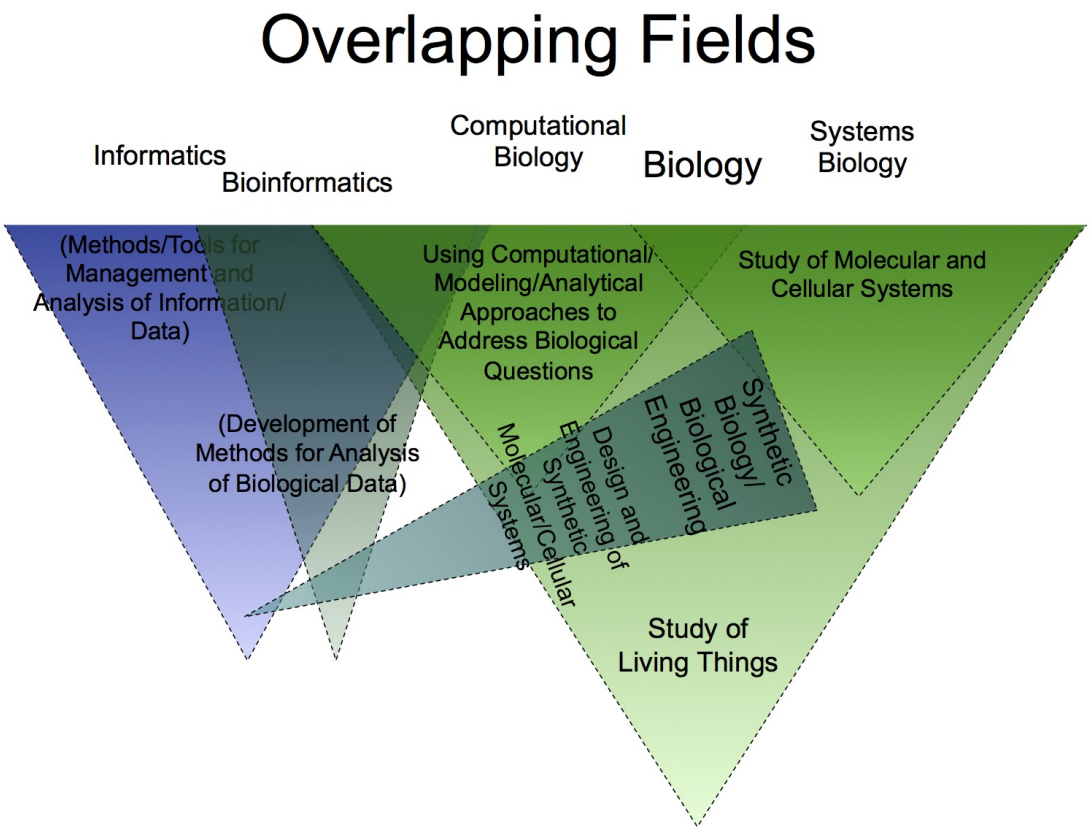

Figure 1.1: A disciplinary map of 'informatics', 'bioinformatics', 'computational biology', 'biology' and 'systems biology'. (Christopher Burge, 'Introduction to Computational and Systems Biology' (7.91), lecture, Massachusetts Institute of Technology, February 8, 2007. Reproduced with permission of Professor Burge.)

of computational biology is more accurately characterised as an evolution rather than a revolution, insofar that it has transformed the modelling of biological systems rather than the understanding of those systems. On the other hand, applying Thomas Kuhn's conception of scientific revolutions, computational biology can be considered a merger of two existing disciplines that was formalised through academic rigour.[10]

To understand this merger, it is necessary to trace the historical origins of both disciplines, starting with the shift away from 'wet lab' or *in vitro* to 'dry lab' or *in silico* research. In the 19th and early 20th century, microbiology research was largely conducted by examining organisms in situ in their environment or collecting and comparing physiological data from different species. This natural history-oriented approach, when combined with insights from geology and physics yielded breakthroughs with respect to species classification and

[10] Thomas S. Kuhn, *The Structure of Scientific Revolutions: 50th Anniversary Edition* (University of Chicago Press, 1972) 6; Rachel A. Ankeny and Sabina Leonelli, 'Repertoires: A post-Kuhnian perspective on scientific change and collaborative research' (2016) 60(Supplement C) *Studies in History and Philosophy of Science Part A* 18–28 22.

phenotypes. The subsequent transition to experimentally-oriented biological research yielded broader applications of biomedical and agricultural research in pharmacology, immunology and industrial fermentation.[11]

As a 2009 report published by the National Research Council (NRC) of the National Academy of Science (NAS) of the US notes, biologists increasingly rely on mathematical and computational models to perform novel analyses of biological data.[12] Jerome Reichman, Tom Dedeurwaerdere and Paul Uhlir note that computational biology tools have the potential significantly improve the research outputs from molecular biology research. These outputs include improving human health and combating pandemics, improving food security by protecting plants and livestock against infection, assisting with conserving biodiversity and addressing energy challenges associated with biofuels.[13] But how did computational tools become so heavily embedded in both basic and applied molecular biology research? And where did the collaborative rules in open source computational biology projects evolve from? To answer these questions, it is necessary to explore the prehistory of molecular biology in the next section of this chapter.

1.2.2 *Sequences and Sequencing - 1945 to 1972*

Although Edna Suarez-Diaz notes that the practices known as 'molecular biology' started to form at the end of the 1930s, it was not until the sequencing work of Frederick Sanger at the University of Cambridge after the end of the Second World War (1939-1945) that the principles underlying molecular biology began to coalesce around an informational approach to protein sequences.[14] Proteins are large molecules, composed of smaller molecules called amino acids and are responsible for most living functions (such as muscular contraction and growth nutrient processing).[15] Examining the evolution of DNA sequencing technology, Miguel Garcia-Sancho argues that the majority of the foundational scientific work underpinning DNA sequencing evolved from protein sequencing.[16] This concept found its genesis within the early sequencing work by Frederick Sanger in examining the chemical structure of the protein insulin. Sanger

---

[11] National Research Council, *A New Biology for the 21st Century* (National Academies Press, 2009) 17; Jerome H. Reichman, Tom Dedeurwaerdere and Paul F. Uhlir, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge University Press, 2016) 8–10.

[12] National Research Council, above n11, 18–19.

[13] Jerome H. Reichman, Tom Dedeurwaerdere and Paul F. Uhlir, above n11, 13-19.

[14] Edna Suárez-Dıaz, 'Making Room for New Faces: Evolution, Genomics and the Growth of Bioinformatics' (2010) 32(1) *History and Philosophy of the Life Sciences* 65–89 67-68.

[15] Miguel Garcıa-Sancho, *Biology, Computing, and the History of Molecular Sequencing: From Proteins to DNA, 1945-2000* (Springer, 2012) 8.

[16] Miguel Garcıa-Sancho, above n15, 6–7.

used the compound dinitrofluorobenzene (DNFB), which reacted with the last amino acids of the insulin chain and dyed them yellow. From this process, Sanger identified that proteins such as insulin had a defined chemical structure.[17]

Sanger's technique for separating and identifying different amino acid chains represented a technological advance in sequencing and identifying the roles of various parts of the insulin protein sequence.[18] Following this breakthrough, other biochemists began attempting to improve the efficiency of protein sequencing.[19] Pehr Edman, a biochemist at the University of Lund, combined Sanger's sequencing technology with paper chromatography to determine the correct sequence order to create a technique known as protein degradation.[20] In turn, Edman's degradation sequencing technique became the basis for the semi-automated sequence determination technology developed by William Stein and Stanford Moore. This sequence determination technology involved successively cutting the amino acids of the protein to identify the missing amino acids by a process of elimination.[21] Crucially for this thesis, the idea of automating sequencing, as well as the length of amino acid sequences, led to the need to adopt quantitative approaches to molecular biology.

Garcia-Sancho notes that the prospect of identifying protein function from DNA or RNA sequences prompted Sanger to shift from protein to genome sequencing.[22] Although Sanger and his colleagues could apply a degradation approach to sequence RNA, the number of sequences in a single DNA molecule made degradation unsuitable.[23] To overcome this limitation, Sanger began to use the enzyme polymerase to make copies of the DNA target region. Sanger used this enzyme to replace the nucleotides in the DNA sequence with chain terminating nucleotides, each labelled with a different width of monochrome radioactive dye. The DNA sequence could then be represented as a diagram called an autoradiograph, which displayed the positions of various nucleotides.[24] The spread of Sanger's technique in molecular biology meant that researchers increasingly directed their enquiries into

---

[17] Frederick Sanger, 'The Free Amino Groups of Insulin' (1945) 39(5) *Biochemical Journal* 507–515; Frederick Sanger, 'The Terminal Peptides of Insulin' (1949) 45(5) *Biochemical Journal* 563–574.

[18] Miguel Garcıa-Sancho, 'A New Insight into Sanger's Development of Sequencing: From Proteins to DNA, 1943–1977' (2010) 43(2) *Journal of the History of Biology* 265–323.

[19] Joel B. Hagen, 'The Origins of Bioinformatics' (2000) 1(3) *Nature Reviews Genetics* 231–236 232.

[20] Pehr Edman and Gregory Begg, 'A Protein Sequenator' (1967) 1(1) *European Journal of Biochemistry* 80–91.

[21] Miguel Garcıa-Sancho, above n18, 309–310.

[22] Miguel Garcıa-Sancho, above n18, 291.

[23] Miguel Garcıa-Sancho, above n18, 297.

[24] Frederick Sanger, 'Sequences, Sequences, and Sequences' (1988) 57(1) *Annual Review of Biochemistry* 1–29; Miguel Garcıa-Sancho, 'Genetic Information in the Age of DNA Sequencing' (2015) 50(1) *Information & Culture* 110–142.

determining the 'sequence of information in the DNA of different organisms'.[25] In turn, this focus led to the increasing integration of quantitative methods in molecular biology.

1.2.3 *The Rise of Quantitative Approaches to Molecular Biology and Genomics*

Joel Hagen argues that the origins of quantitative approaches in molecular biology can be found in the efforts by natural scientists to systematise taxonomy and to solve problems of population and geographical distribution.[26] Edna Suarez-Diaz and Victor Anaya-Munoz posit that the search for statistical models to represent naturally occurring populations was driven by the desire to transform natural biology from a subjective science into an objectively measured science and to assume 'a statistical frame of mind'.[27] In other words, by shifting away from manual, subjective comparisons towards statistical comparisons of the characteristics of different organisms, biologists could produce more easily repeatable research.[28] As Hagen notes, theoretical biologists and natural scientists quickly realised that species clustering and other statistical techniques were extremely cumbersome to carry out by hand.[29] The emergence of fast digital computers meant that statistical techniques were increasingly accessible to molecular biologists in both academic and industrial research environments.

1.3 THE EVOLUTION OF COMPUTER SCIENCE AND SOFTWARE ENGINEERING

1.3.1 *The Evolution of Programming Languages - 1954 to 1960*

Hennessy and Patterson define computer hardware as 'the [physical] specifics of a computer, including the detailed logic design and packaging technology'.[30] Hardware can be contrasted with the encoded information that runs on these physical platforms, which is collectively referred to as software.[31] This distinction first emerged with the development of formal

---

[25] Miguel Garcıa-Sancho, above n18, 311.

[26] Joel B Hagen, 'The Introduction of Computers into Systematic Research in the United States During the 1960s' (2001) 32(2) *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 291–314 291–292.

[27] Edna Suárez-Dıaz and Victor H. Anaya-Muñoz, 'History, Objectivity, and the Construction of Molecular Phylogenies' (2008) 39(4) *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 451–468.

[28] Joel Hagen, 'The Statistical Frame of Mind in Systematic Biology from Quantitative Zoology to Biometry' (2003) 36(2) *Journal of the History of Biology* 353–384 377.

[29] Joel B. Hagen, above n19, 234; Joel B Hagen, above n26 291–292.

[30] John L. Hennessy and David A. Patterson, *Computer Architecture: A Quantitative Approach* (Elsevier, 2012) 15.

[31] J Glenn Brookshear, *Computer Science: An Overview* (Pearson/Addison Wesley, 2009) 2.

programming languages.[32] Prior to 1954 programmers wrote instructions for digital computers in machine code.[33] Machine code operates by literally providing the addresses for memory locations and transistor registries that send and store data.[34] Machine code represents a highly efficient way of programming (and is still used today for certain tasks, such as programming embedded hardware components). However, the process for removing errors is a time consuming process, particularly for programmers unfamiliar with the hardware circuit design for a particular platform.[35]

The next phase in the evolution of programming languages was the development of assembly languages, which allowed programmers to refer to machine code functions with mnemonics and memory addresses with symbolic labels.[36] However, the syntax for assembly languages still relied on hardware specific notations, which required reprogramming for translation to different hardware platforms and would be difficult for subsequent programmers to understand.[37] As a result, computer scientists designed high level languages (HLLs) that further abstracted functions being performed by hardware.[38] In addition to being hardware independent, all of these HLLs could be used to produce software that operated functionally rather than procedurally.[39] In other words, rather than performing a series of tasks in a sequential order, HLLs allowed programmers to call on certain subroutines in a non-sequential order.[40]

As high-level programming languages decreased the amount of coding required for software, computer programs began to increase in complexity. Early HLLs (particularly LISP) included subroutines as a stylistic choice to encourage function-oriented programming. Nevertheless, programmers began to use subroutines to coherently structure code, as well as

---

[32] James A. Whittaker and Jeffrey M. Voas, '50 years of software: key principles for quality' (2002) 4(6) *IT Professional* 28–35; J Glenn Brookshear, above n31, 77.

[33] J Glenn Brookshear, above n31 77; Thomas Haigh and Mark Priestley, 'Where Code Comes From: Architectures of Automatic Control from Babbage to Algol' (2016) 59(1) *Communications of the ACM* 39–44 40-41.

[34] J Glenn Brookshear, above n31, 240.

[35] J Glenn Brookshear, above n31 240; Frank Vahid and Tony Givargis, *Embedded System Design: a Unified Hardware/Software Introduction* (Wiley, 2002).

[36] Joey Paquet and Serguei A. Mokhov, 'Comparative Studies of Programming Languages; Course Lecture Notes' (2010) *arXiv:1007.2123 [cs]*.

[37] J Glenn Brookshear, above n31, 241–2; Justin Joque, 'The Invention of the Object: Object Orientation and the Philosophical Development of Programming Languages of Programming Languages' (2016) 29(4) *Philosophy & Technology* 335–356 340.

[38] Justin Joque, above n37, 341.

[39] Marc Nerlove, 'Programming Languages: A Short History for Economists' (2004) 29(1-3) *Journal of Economic & Social Measurement* 189–203 193.

[40] Jean E. Sammet, *Programming Languages: History and Fundamentals* (Prentice-Hall, Inc., 1969) 14-19.

rely on modular programming techniques.[41] This modular approach to programming has a number of advantages compared to sequential programming. As Marc Nerlove notes, John McCarthy's efforts with LISP to separate functionality into modules had an enormous impact on the development of object-oriented programming.[42] As the next subsection of this Chapter will discuss, the emergence of new object-oriented languages played a fundamental role in the development of modern data types, software libraries and time sharing operating systems. These developments in turn supported the cyberinfrastructure which itself supported early forays into computational biology discussed in Section 1.3.4.[43]

To articulate these concepts further, it is necessary to explain what object-oriented programming entails. At the software language level, modularity is implemented through inheritance, where new data types (or representations of digits and numbers) are defined as extensions of previous data types (called 'classes').[44] Each of these classes may also inherit particular functions from previous classes, known as methods. The benefits of inheritance are that, rather than having to repeat code, a programmer can use classes, methods and data types developed by other programmers.[45] Individual components can be developed in isolation and then built into a larger system.[46] This design approach led to developers collecting methods together to form libraries, following the shift toward software engineering in 1968. Libraries range from providing routine programming functions such as variable manipulation or creating certain data types, to providing specific functions such as scientific operations.[47]

Although libraries were originally designed for programmers working together internally in an institution, there has been an increasing trend to make libraries available online. To ensure the portability and reusability of software, these libraries can be called from other programs using Application Programming Interfaces (APIs). Modern software systems are commonly composed of a large set of applications and libraries which are developed by separate developers in isolation from one another.[48] In other words, an object-oriented

---

[41] Peter Naur and Brian Randell, 'Software Engineering' (Report of a Conference Sponsored by the NATO Science Committee, NATO Scientific Affairs Division, Garmisch, 7-11 October 1968)

[42] Marc Nerlove, above n39, 196.

[43] P. Umesh et al., 'Programming Languages for Synthetic Biology' (2010) 4(4) *Systems and Synthetic Biology* 265–269 265.

[44] Stephen R. Schach, *Classical and object-oriented software engineering with UML and C++* (WCB/McGraw-Hill, 1999) 201–202.

[45] Stephen R. Schach, above n44 201–202.

[46] David L. Parnas, 'Designing Software for Ease of Extension and Contraction' (Paper presented at *Proceedings of the 3rd International Conference on Software Engineering* , 1978).

[47] John. R. Rice and Ronald. F. Boisvert, 'From Scientific Software Libraries to Problem-Solving Environments' (1996) 3(3) *IEEE Computational Science and Engineering* 44–53 45.

[48] Jesus M. Gonzalez-Barahona et al., 'Macro-level Software Evolution: A Case Study of a Large Software Compilation' (2009) 14(3) *Empirical Software Engineering* 262–285 280.

approach allows software to be developed in a modular way that assists multiple different developers in contributing to build a single software system, rather than one developer or development team taking full responsibility for development. This tendency is inherent in the history of modern operating system development, which is discussed in the next section of this chapter.

1.3.2 *Operating Systems and Cyberinfrastructure - 1969 onwards*

Sections 1.3.1 to 1.3.3 noted the divergence in the development of hardware and software, and in particular the shift towards platform independent software development. This shift was accompanied by the development of operating systems software. An operating system is a specialised program which manages computer hardware resource allocation to other programs.[49] With the early digital computers, it was only possible to run one process at any one time, so programmers developed time sharing systems to allow multiple users to connect to one computer. These developments were coupled with upgrades to random access memory (RAM) which meant that more than one program could be stored and run at any one time (otherwise known as multiprocessing).[50]

The first multiprocessing operating system, MULTICS, was developed by Bell Laboratories in 1965. Bell Laboratories (a research subsidiary of American Telephone Telegraph (AT&T)) was prohibited from selling software. This prohibition was eventually formalised by a consent decree imposed by the US Department of Justice in 1956 requiring AT&T to divest the local exchange operations. This consent decree also prohibited AT&T from marketing advanced mechanical or electrical engineering components beyond regulated carrier services.[51] Ken Thompson and Dennis Ritchie (the original inventors of MULTICS) started work on developing a portable version of MULTICS, called UNIX, written in C, an object-oriented programming language. Thompson and Ritchie distributed the UNIX source code to interested parties, at first by post and only at the cost of postage and storage media.[52] In the absence of formal licensing, Thompson and Ritchie developed their own informal norm-based system for introducing changes back into the main UNIX operating system (or kernel). Particular emphasis was placed on recipients of UNIX being able to freely modify and share the UNIX system. Significant improvements to UNIX flowed from this development approach.

This release pattern remained relatively undisturbed until 1976, when the terms of the

---

[49] J Glenn Brookshear, above n31, 110.

[50] Andrew S. Tanenbaum, *Modern Operating Systems* (Pearson Prentice Hall, 2008) 6–8.

[51] C. L. Brown, 'AT&T and the Consent Decree' (1983) 7(2) *Telecommunications Policy* 91–95 94.

[52] Dennis M. Ritchie, 'Reflections on Software Research' (2012) 17(8) *Resonance* 810–816 812.

consent decree imposed by the Department of Justice in 1956 were renegotiated to allow AT&T to licence computer programs. This renegotiation permitted AT&T to sell a proprietary version of UNIX for 300 US dollars.[53] The subsequent terms upon which AT&T distributed UNIX were highly restrictive, preventing licensees from writing technical manuals or teaching UNIX coding to students.[54] However, Thompson and Ritchie's earlier distribution network for UNIX meant that there were multiple independent implementations of UNIX compatible with AT&T's minimum UNIX hardware compatibility standards.[55] These UNIX distributions included several 'open source' or 'free software' implementations (such as the Berkeley Software Distribution (BSD) developed at University of California Berkeley and MINIX developed at the Vrije Universiteit Amsterdam). Because there were no licence fees associated with these open source implementations, and they were capable of running on a wide variety of hardware, they became extremely popular in educational institutes and in scientific research.[56] The open source versions of UNIX demonstrate how the underlying object-oriented language used in UNIX can support software source code sharing.[57]

This collaborative development model was also aided by the early development and expansion of DARPANet (which was run from the US Department of Defence Advanced Research Projects Agency (ARPA)). First developed in 1969 in collaboration with the private telecommunications firm Bolt, Beranek and Newman (BBN), ARPANet was one of the first attempts to build a decentralised-computer to-computer network that was reliant on 'packet switching'.[58] Packet switching involves transferring information over a network using headers to describe what each packet of information contains. Early research into packet switching was driven by an interest in developing an alternative to centralised phone switches that would not be as vulnerable in the event of nuclear war.[59] Accordingly, the development of the Network Control Protocol (NCP) and later Transmission Control Protocol (TCP/IP) as part of the DARPANet project indubitably represented an important shift towards standardised

---

[53] Andrew S. Tanenbaum, 'Lessons Learned from 30 Years of MINIX' (2016) 59(3) *Communications of the ACM* 70–78 70.

[54] Andrew S. Tanenbaum, above n53, 71.

[55] Joel West, 'How Open is Open Enough?: Melding Proprietary and Open Source Platform Strategies' (2003) 32(7) *Research Policy* 1259–1285 1263.

[56] Michael Schwarz and Yuri Takhteyev, 'Half a Century of Public Software Institutions: Open Source as a Solution to Hold-Up Problem' (2010) 12(4) *Journal of Public Economic Theory* 609–639 623–624.

[57] Peter H. Salus, *A Quarter Century of UNIX* (Addison-Wesley Publishing Company, 1994) 70; Johan Söderberg, *Hacking Capitalism: The Free and Open Source Software Movement* (Routledge, 2015).

[58] Andrew L. Russell and Valérie Schafer, 'In the Shadow of ARPANET and Internet: Louis Pouzin and the Cyclades Network in the 1970s' (2014) 55(4) *Technology and Culture* 880–907 882.

[59] Peter J. Denning, 'The Science of Computing: The ARPANET after Twenty Years' (1989) 77(6) *American Scientist* 530–534 533.

network communication.[60]

The open TCP/IP standard was incorporated into the majority of UNIX distributions (most notably the Berkeley Software Distribution (BSD) produced by the University of Berkeley.[61] The implementation of the TCP/IP standard allowed developers of UNIX to connect their machines with one another, as well as exchange software source code.[62] Standardised network protocols for file sharing were particularly important in academic environments. In addition, prior to the establishment of restrictive licensing conditions, UNIX and its derivatives had become the predominant platform for academic and scientific computing due to its portability.[63] As the later sections of this chapter will explore, ARPANet (and equivalent research networks) also played a vital role in helping to establish early collaborative scientific research initiatives by providing a mechanism for researchers to share data.[64] The next section considers the important role that both open operating system development and network technology played in the expansion of early computational biology initiatives.

1.4 THE MERGER OF MOLECULAR BIOLOGY AND COMPUTER SCIENCE - THE ORIGINS OF COMPUTATIONAL BIOLOGY

Chow-White and Garcia-Sancho argue that the merger between molecular biology and computer science was best reflected in three key areas of convergence. These three areas of convergence are sequencing automation and recombinant DNA, sequencing databases and sequencing algorithms.[65] Although these developments occurred simultaneously, they each had a significant technical impact on improvements on one another. Sections 1.4.1 to 1.4.5 will address each of these developments respectively.

1.4.1 *Automating Genome Sequencing and Recombinant DNA*

Whilst Sanger focused on more accurate sequencing techniques, other research teams focused their attention on complete automation of the sequencing process, particularly the replication of DNA sequences. The research team led by Leroy Hood at the California Institute of

---

[60] A. Michael Froomkin, 'Habermas@Discourse.Net: Toward a Critical Theory of Cyberspace' (2002) 116(3) *Harvard Law Review* 749–873 778.

[61] John S. Quarterman, Abraham Silberschatz and James L. Peterson, '4.2BSD and 4.3BSD As Examples of the UNIX System' (1985) 17(4) *ACM Computing Surveys* 379–418.

[62] Peter H. Salus, above n57, 164.

[63] Dennis M. Ritchie, above n52 812.

[64] Dennis M. Jennings et al., 'Computer Networking for Scientists' (1986) 231(4741) *Science* 943–950.

[65] Peter A. Chow-White and Miguel Garcıa-Sancho, 'Bidirectional Shaping and Spaces of Convergence: Interactions between Biology and Computing from the First DNA Sequencers to Global Genome Databases' (2012) 37(1) *Science, Technology, & Human Values* 124–164 125.

Technology ('Caltech') sought to remove human intervention from the process of identifying different nucleotides in the autoradiographs.[66] Whilst it was relatively easy to manually distinguish between different segments in the autoradiograph, the automation technology at the time could not distinguish between blots of the same shape and colour.[67] Rather than using a monochrome dye, Hood's team changed direction and concentrated on using different coloured dyes to distinguish between different nucleotide bases (for example, red for guanine). Each of these colours could be represented as a different character in the computer attached to the sequencer once scanned via laser. This approach made it easier for pattern identification software embedded in a sequencer to process each nucleotide within a particular sequence. This approach to automated sequencing cemented an informational approach to molecular biology where genotypic information could be used to predict phenotypic characteristics.[68]

Operating simultaneously to Hood's team, Stanley Cohen and Herbert Boyer of Stanford University had begun research into recombinant DNA (or 'rDNA') technology to create sequences that would not otherwise be found within a genome.[69] Cohen and Boyer first contextualised rDNA technology as a mechanism for identifying the functioning of specific mammalian genes by inserting these genes into the chromosomes of bacteria such as *Escherichia coli* (*E. Coli*). However, together with Paul Berg (also of Stanford), they soon realised that rDNA technology could also be used to modify the genomes of existing organisms.[70] From a theoretical perspective, rDNA technology also paved the way for further research in engineering or synthetic biology towards editing organism genotypes.[71]

Automated sequencers and recombinant DNA also dramatically increased the use of sequencing in molecular biology laboratories. The Caltech research team sought to commercialise this automated sequencing technology so that they could sell it to academic research environments. Peter Chow-White and Miguel Garcia-Sancho attribute this emphasis on commercialisation to Caltech's constitution, which permitted it to seek funding through research contracts.[72]. This emphasis on commercialisation can be seen from Hood's team

---

[66] Lloyd M. Smith et al., 'Fluorescence Detection in Automated DNA Sequence Analysis' (1986) 321(6071) *Nature* 674–679.

[67] Peter A. Chow-White and Miguel Garcıa-Sancho, above n65, 133.

[68] Miguel Garcıa-Sancho, above n24, 135.

[69] Miguel Garcıa-Sancho, above n24, 112–113.

[70] Paul Berg and Janet E. Mertz, 'Personal Reflections on the Origins and Emergence of Recombinant DNA Technology' (2010) 184(1) *Genetics* 9–17 10–11.

[71] Institute of Medicine (US) Forum on Microbial Threats, *The Science and Applications of Synthetic and Systems Biology: Workshop Summary* (National Academies Press (US), 2011) 1–2; Maarten Boudry and Massimo Pigliucci, 'The Mismeasure of Machine: Synthetic Biology and the Trouble with Engineering Metaphors' (2013) 44(4) *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 660–668 666.

[72] Peter A. Chow-White and Miguel Garcıa-Sancho, above n65 133.

capitalising on this development and establishing a spin off company, Applied Biosystems, to market the first automated sequencers.[73] Likewise, Boyer and Cohen successfully commercialised their recombinant DNA methods through patent protection, albeit relying on a non-exclusive licensing model.[74] Of particular relevance to this thesis was the sudden flood in the amount of genomic sequence data that molecular biologists were producing as a result of the spread of sequencing technology.[75] This flood of sequencing data was accompanied by significant data sharing initiatives between different research institutes.

1.4.2 *The 'Atlas of Protein Sequence and Structure' - The First Sequence Compilations and Bioinformatics Algorithms*

Hagen notes that advances within sequencing techniques (along with the promise of automated sequencing) encouraged laboratories to collect the information extracted from protein sequences into libraries. In 1960, Margaret Dayhoff was appointed by the newly established National Biomedicine Research Foundation (NBRF) to explore mathematical approaches for analysing amino acid sequence data.[76] Dayhoff, along with Robert Ledley and Richard Eck, began writing software in FORTRAN to determine maps of possible amino-acid sequences. Dayhoff, Ledley and Eck's programs were able to determine the correct sequence structure within a matter of minutes, as opposed to the months that it would take to manually sequence by hand.[77] Combined with automatic amino acid sequences, such as those developed by Edman, Stein and Moore, Dayhoff quickly developed a library of sequences that could be used for comparative biochemistry and molecular evolution study.[78]

In 1965, Dayhoff, Ledley and Eck compiled and catalogued all of their reconstructed protein sequences into an annual publication, The Atlas of Protein Sequence and Structure ('the Atlas') as an aid to other molecular biology researchers.[79] Bruno Strasser notes that Dayhoff, Ledley and Eck were not the only research team providing a compilation of

---

[73] Peter A. Chow-White and Miguel Garcıa-Sancho, above n65, 134–135.

[74] Martin Kenney and Donald Patton, 'Reconsidering the Bayh-Dole Act and the Current University Invention Ownership Model' (2009) 38(9) *Research Policy* 1407–1422 1417-8.

[75] Rachel A. Ankeny, 'The Natural History of Caenorhabditis elegans Research' (2001) 2(6) *Nature Reviews Genetics* 474–479; A. J. G. Hey and A. E. Trefethen, 'The Data Deluge: An e-Science Perspective' in F. Berman, G. C. Fox and A. J. G. Hey (eds.), *Grid Computing - Making the Global Infrastructure a Reality* (Wiley and Sons 2003)809–824 3.

[76] Joel B. Hagen, above n19 234.

[77] Margaret Oakley Dayhoff and Robert S. Ledley, 'Comprotein: A Computer Program to Aid Primary Protein Structure Determination' (Paper presented at *Proceedings of the December 4-6, 1962, Fall Joint Computer Conference* , 1962).

[78] Joel B. Hagen, above n19 234.

[79] Margaret Oakley Dayhoff and National Biomedical Research Foundation, *Atlas of Protein Sequence and Structure. [Vol. 1], [Vol. 1],* (National Biomedical Research Foundation, 1969).

sequence data. For example, Olga Kennard at the University of Cambridge established the Cambridge Crystallographic Data Centre in 1965 as a repository of data on the molecular structure of organic chemicals, and the Protein Data Bank was established in 1973 and hosted at Brookhaven National Laboratory as a repository of the atomic co-ordinates of protein structures. Likewise, the European Organisation for Nuclear Research (CERN) provided the basis for the establishment of the European Molecular Biology Organisation (EMBO). EMBO was founded as a supranational molecular biology research network for the European Union, and later formed the core of the European Molecular Biology Laboratory (EMBL).

However, Dayhoff and her team were the first to make their database available, initially on a 'gratis' basis to assist other researchers.[80] Dayhoff, Ledley and Eck also encouraged other researchers to contribute their own sequences to the database in exchange for access to the rest of the database. Strasser refers to this system of distribution as a 'network of exchange' and a 'gift based economy'.[81] The politics of this early sequencing movement and the development of sequence databases such as the Atlas also prompted further research into algorithms to more efficiently and objectively compare sequence samples without the need to rely on manual examination.[82]

A significant breakthrough occurred when Walter Fitch and Emanuel Margoliash developed an algorithm which could be used to calculate the mutation distances. Mutation distances referred to the minimum number of steps required to convert one amino acid chain in a protein sequence to another. These mutation distances then could be constructed into phylogenetic trees, or tree diagrams that demonstrated the inferred evolutionary relationships amongst organisms based on the mutation distances between each sequence.[83] Following simultaneous algorithmic research by Dayhoff, Ledley and Eck, the *Atlas* began to include phylogenetic trees of common model organisms based on their evolutionary characteristics.[84] These phylogenetic trees not only allowed other researchers compare certain 'model' organisms, but also assisted experimental biologists speed up their research by allowing them to compare protein functionality across a few model organisms.[85] As Section 1.2 discussed,

---

[80] Bruno J. Strasser, 'The Experimenter's Museum: GenBank, Natural History, and the Moral Economies of Biomedicine' (2011) 102(1) *Isis* 60–96 67.

[81] Bruno J. Strasser, 'Collecting, Comparing, and Computing Sequences: The Making of Margaret O. Dayhoff's Atlas of Protein Sequence and Structure, 1954–1965' (2009) 43(4) *Journal of the History of Biology* 623–660 644.

[82] Jeffrey L. Thorne, Hirohisa Kishino and Joseph Felsenstein, 'An Evolutionary Model for Maximum Likelihood Alignment of DNA Sequences' (1991) 33(2) *Journal of Molecular Evolution* 114–124.

[83] Walter M. Fitch and Emanuel Margoliash, 'Construction of phylogenetic trees' (1967) 155(3760) *Science (New York, N.Y.)* 279–284.

[84] Bruno J. Strasser, above n81, 624.

[85] Edna Suárez-Dıaz and Victor H. Anaya-Muñoz, above n27, 455.

the goal of improving algorithm efficiency was largely driven by the perceived need to ensure the efficiency and transparency of computational methods in molecular biology.[86] This goal reflects the importance of reproducibility of research results in computationally driven research.[87] As Section 1.5 of this chapter reveals, the need to ensure reproducibility is one of the predominant justifications for open licensing of scientific software and data.

1.4.3 *Networked Databases and Bioinformatics - from the Atlas to GenBank*

The Atlas was extremely popular amongst molecular biology researchers involved in natural history research. However, its editors soon encountered problems with respect to the experimental molecular biology community. Strasser notes that Dayhoff's decision not to prioritise or provide individual attribution to authors, which was the norm within natural history research at the time, created significant antagonism amongst experimental researchers over ownership of methods.[88] As a result of this hostility, many experimental molecular biologists refused to contribute to the Atlas. Dayhoff and her team were forced to scour the literature and pay copyright royalties to include existing published sequences outside of the work that her team had performed. To gain access to these publications, Dayhoff was forced to charge a subscription fee from subscribers to the Atlas (as well as postage fees for the electronic tapes which the Atlas was stored on). Dayhoff justified the decision to seek reimbursement on the grounds that she and her team could not otherwise cover the expenses of publishing it. Dayhoff charged 400 US dollars for magnetic tapes from the Atlas (relative to 35 US dollars from the Protein Data Bank), further deepening the distrust towards the Atlas amongst the experimental molecular biology community.[89]

The death blow to the Atlas came when the National Institutes of Health (NIH) issued a request for proposals on developing a nucleic acid sequence computerised database. The need for this database was exacerbated by the release of the publicly available sequence database first offered by EMBL in 1980.[90] The two competing contracts were from Dayhoff's team at NBRF and a group headed by Walter Goad at the Los Alamos National Laboratory in a private collaboration with BBN, who had already played an important role in the development of DARPANet.[91] Perhaps somewhat cynically, Goad's application stressed that, unlike Dayhoff's

---

[86] Edna Suárez-Dıaz and Victor H. Anaya-Muñoz, above n27, 456.

[87] Victoria Stodden et al., 'Enhancing Reproducibility for Computational Methods' (2016) 354(6317) *Science* 1240–1241.

[88] Bruno J. Strasser, above n80, 81.

[89] Hallam Stevens, *Life Out of Sequence: A Data-Driven History of Bioinformatics* (University of Chicago Press, 2013) 146.

[90] Peter A. Chow-White and Miguel Garcıa-Sancho, above n65, 133.

[91] Bruno J. Strasser, 'Genetics. GenBank–Natural history in the 21st Century?' (2008) 322(5901) *Science (New York,*

team, the Los Alamos National Laboratory-BBN partnership would not 'assert any proprietary interests whatsoever in any data'. Implicitly, this application contrasted this pledge with how Dayhoff had already sought reimbursement for the ongoing publication of the Atlas. Instead, experimental molecular biologists would submit their own data to the database along with annotations so as to maintain their attribution rights.

The NIH therefore signed a contract with the Los Alamos National Laboratory-BBN partnership to establish a networked molecular sequence database that became known as GenBank.[92] Dayhoff and the NBRF soon after ceased publishing the Atlas. Dayhoff and her team nevertheless paved the way for both large compilations of sequence data and algorithms to quantitatively analyse protein, RNA and DNA sequences. The fight for the GenBank contract is demonstrative of a number of important principles about computational biology, publicly funded research and the role of intellectual property rights. Goad was able to convince the NIH that the Los Alamos National Laboratory would not assert any proprietary right over any sequence data or sequencing software associated with GenBank.

However, there was no formal funding model attached to the Los Alamos National Laboratory proposal at the time it was submitted. Further, it took three years for an appropriate funding model to emerge from the NIH.[93] Whilst the discussion in this section on objectivity, repeatability and reproducibility highlights the importance of open source development practices in scientific software research, it is questionable whether open source licensing will succeed without an ongoing plan for sustainability. Indeed, it is this question which is central to the second and third research questions posed by this thesis. The next section will concentrate on the HGP and the International HapMap Project, not only as the first examples of large scale networked computational biology initiatives, but also as computational biology initiatives with co-ordinated strategies towards open innovation.

1.4.4 *From Sequence Databases to Large Scale Computational Biology - The Role of Computational Biology in the Human Genome Project and the International HapMap Project*

The Los Alamos National Laboratory's success in obtaining the contract to run the GenBank database was followed by two key developments. The first was increasing integration of sequence collections between GenBank and the two other major sequence collections: the EMBL sequence collection; and the DNA Database Bank of Japan (DDBJ) as part of a

---

N.Y.) 537–538.

92 Bruno J. Strasser, above n91, 538.

93 Teresa. K. Attwood et al., 'Concepts, historical milestones and the central place of bioinformatics in modern biology: a European perspective' in *Bioinformatics - trends and methodologies* (InTech 2011) 11.

collaboration initiated by Goad. As traditional taxonomic labelling strategies were no longer acceptable, the increased integration and sharing of data encouraged shared data standards. There was still residual resistance amongst the experimental biology community towards the concept of data sharing. However, Strasser notes that established sequence data standards encouraged some journal editors to adopt the same standards and mandate electronic submission of data.[94]

Coupled with the socio-technical shift towards computational biology practices was the increasing recognition of the applied potential of informational approaches to molecular biology, particularly in human healthcare management.[95] Garcia-Sancho notes that Sanger, Walter Gilbert and Leroy Hood argued that sequence databases and maps could be used to identify the function of human genes and shift biomedicine from a reactive to a preventative model.[96] Consequently, in 1984 the US Department of Energy began to explore the idea of establishing a large scale project for mapping the human genome.[97] This target was highly ambitious. In 1984, the most advanced of Hood's automated sequencers could only read 300 base pairs in a single analysis and the human genome contained over 3 billion base pairs. From its inception, the HGP was therefore highly reliant on a large number of sequencers working in concert to complete the project as well as the software infrastructure that had been established through GenBank.

In addition, one of the key factors in the NIH granting the Los Alamos Laboratory the contract to establish a nucleic acid sequence database was the computational and network infrastructure that Goad and his team had at their disposal. One of the nascent computer science laboratories that took receipt of Thompson and Ritchie's original UNIX distributions was the Stanford Artificial Intelligence Laboratory (SAIL).[98] SAIL was established as part of Stanford University's initiative to offer an expert system for medical diagnostic services remotely out of Stanford University. Specifically, SAIL provided the networking hardware support for the Stanford University Medical Experimental Computer for Artificial Intelligence in Medicine (SUMEX-AIM).[99]

However, Joshua Lederberg, the creator of SUMEX-AIM, saw SAIL's potential as a

---

[94] Bruno J. Strasser, above n91 538.

[95] National Research Council, above n11 18–19.

[96] Miguel Garcıa-Sancho, above n24, 132–133.

[97] Robert M. Cook-Deegan, *The Gene Wars: Science, Politics, and the Human Genome* (W.W. Norton & Company, 1996).

[98] Hallam Stevens, 'Networking Biology: The Origins of Sequence-Sharing Practices in Genomics' (2015) 56(4) *Technology and Culture* 839–867 844.

[99] Elliott C. Levinthal et al., 'When Computers Talk To Computers' (1975) 17(12) *Industrial Research* 35–42; Joseph November, 'Removing the Center from Computing: Biology's New Mode of Digital Knowledge Production' (2011) 34(2) *Berichte zur Wissenschaftsgeschichte* 156–173 170.

distributed network system to share biomedical data between American research institutes and universities.[100] In particular, Lederberg's perspectives on the open exchange of biomedical and molecular biology data as a solution to mitigate against the negative effects of patents on basic molecular biology research were also influential.[101] Accordingly, in 1975 Lederberg used the networking infrastructure from SUMEX-AIM to establish MOLGEN.[102] MOGLEN was not designed to provide remote hospitals with access to Stanford's diagnostic facilities. Instead, the aim of the MOLGEN project was to provide molecular biologists across the United States with network access to artificial intelligence and expert system tools for use in molecular biology research.[103]

Eventually, the SUMEX-AIM committee approved the creation of GENET, a networking platform which allowed molecular biologists to login to the SUMEX-AIM network and not only share sequence data but also use particular sequencing tools.[104] In addition, SUMEX-AIM provided molecular biologists with a gateway to access GenBank's nucleic acid sequence collection. GENET proved so popular amongst the molecular biology research community that it overwhelmed the available processing power on SUMEX-AIM. However, Peter Frieldland and Lawrence Kedes, two graduate students at Stanford, subsequently established a private company, Intelligenetics, to run a version of GENET called BIONET.[105] The BIONET system allowed a wide range of molecular biologists, not just those based at Stanford University, to benefit from access to GenBank's sequencing collections.[106]

Moreover, following the establishment of the HGP, BIONET and GenBank played a crucial role in providing the underlying infrastructure as a means for participants to share and deposit human genome sequence data respectively.[107] Initially, the six key partner nations to the HGP

---

[100] Joshua Lederberg, 'Digital Communications and the Conduct of Science: The New Literacy' (1978) 66(11) *Proceedings of the IEEE* 1314–1319.

[101] Doogab Yi, 'Cancer, Viruses, and Mass Migration: Paul Berg's Venture into Eukaryotic Biology and the Advent of Recombinant DNA Research and Technology, 1967–1980' (2008) 41(4) *Journal of the History of Biology* 589–636 626; Hallam Stevens, above n98, 851.

[102] MOLGEN was officially known as Applications of Symbolic Computation and Artificial Intelligence to Molecular Biology

[103] Hallam Stevens, above n98, 851–2.

[104] Hallam Stevens, above n98, 852–3.

[105] Dennis H. Smith et al., 'BIONET: National Computer Resource for Molecular Biology' (1986) 14(1) *Nucleic Acids Research* 17–20.

[106] Hallam Stevens, above n98, 862–3; Charles Lawson and Michelle Rourke, 'Open Access DNA, RNA and Amino Acid Sequences: The Consequences and Solutions for the International Regulation of Access and Benefit Sharing' (2016) 24(3) *Journal of Law and Medicine* 96–118.

[107] Dianne Nicol, 'Public Trust, Intellectual Property and Human Genetic Databanks: the Need to Take Benefit Sharing Seriously' (2006) 3(3) *Journal of International Biotechnology Law* 89–103 91; Robert Cook-Deegan, Rachel A. Ankeny and Kathryn Maxson Jones, 'Sharing Data to Build a Medical Information Commons: From Bermuda to the Global Alliance' (2017) 18(1) *Annual Review of Genomics and Human Genetics* 389 7.

were the US, France, Germany, Japan, China and the United Kingdom. Of particular relevance to this thesis is the role that the networking databases played in the completion of the HGP. In particular, the interconnected GenBank, EMBL and DDBJ allowed researchers from non key partner nations to participate in the HGP. Australia and New Zealand, despite relatively slow network speeds compared to the six original partner nations, were able to make substantial contributions to the HGP.[108]

BIONET is a good example of how allowing molecular biologists with a networked computer to access and deposit sequence samples could help drive large scale sequencing projects.[109] In particular, Garcia-Sancho and Cook-Deegan argue that large scale scientific projects such as the HGP often start as a project run by smaller institutions. By relying on networking technology, these smaller projects can grow beyond their institutional and geographic constraints.[110] This concept of institutional organisation and its impact on both open source licensing and data sharing within computational biology research will be covered in greater detail in Chapter Four. The question of institutional organisation is in turn related to the commercialisation of computational biology software and data. At the time BIONET was established, there was unease amongst the molecular biology community about the prospect of IntelliGenetics, an ostensibly private company, receiving government funding to oversee the BIONET and GenBank network infrastructure.[111] Section 1.5, as well as later chapters of this thesis, address the issue of how to balance government funding of commercial software against the need for reproducibility.

As Cook-Deegan notes, the debate over software ownership, funding and commercialisation foreshadowed efforts to commercialise the sequencing of the human genome.[112] J. Craig Venter, an NIH scientist working with Leroy Hood on automated genome sequencing, started to collect thousands of fragments of DNA from brain tissue. Venter then identified the parts of these sequences (called express sequence tags or ESTs) that corresponded to certain proteins responsible for forming brain tissue.[113] However, there was a firestorm of controversy when Venter (as an employee of the NIH) filed for patents on

---

[108] Mark A. Ragan, Tim Littlejohn and Bruce Ross, 'Genome-Scale Computational Biology and Bioinformatics in Australia' (2008) 4(8) *PLOS Comput Biol* e1000068 2.

[109] Francis S. Collins, Michael Morgan and Aristides Patrinos, 'The Human Genome Project: Lessons from Large-Scale Biology' (2003) 300(5617) *Science* 286–290 289.

[110] Miguel Garcıa-Sancho, 'The Proactive Historian: Methodological Opportunities Presented by The New Archives Documenting Genomics' (2016) 55(February) *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 70–82; Robert Cook-Deegan, Rachel A. Ankeny and Kathryn Maxson Jones, above n107.

[111] Leslie Roberts, 'The Perils of Involving Congress in a 'Catfight'' (1992) 257(5067) *Science* 156–157; Robert M. Cook-Deegan, above n97, 291–292; Hallam Stevens, above n98, 854.

[112] Robert M. Cook-Deegan, above n97, 292.

[113] Robert M. Cook-Deegan, above n97, 314.

ESTs.[114] Venter justified the decision on the grounds that the US Supreme Court had declared a decade earlier that genetically modified organisms could be patented. without this protection the NIH could be circumvented by a privately owned Japanese or European firm.[115] Nevertheless, after immense internal and external pressure surrounding concern about EST patents stifling collaborative research the NIH abandoned their EST patent applications.[116] In response, Venter founded his own non-profit research institute, The Institute for Genome Research (TIGR). TIGR then assigned EST patents to Human Genome Services (HGS), a for profit company.[117]

From this point, there was a race between the government funded HGP and the privately funded TIGR-HGS group to sequence the human genome.[118] The TIGR-HGS consortium (later reconsolidated into Celera Genomics) initially outflanked the HGP by relying on the faster shotgun sequencing method. This method utilised proprietary algorithms to shred and then reconstruct a whole sequence from the fragments. To bypass the potential for the TIGR-HGS consortium to apply for patents, the HGP consortium established the Bermuda Principles in 1996. The Bermuda Principles mandated that all member states place human genomic sequence data onto GenBank nodes as soon as possible after sequencing. However, the public availability of the GenBank results meant that TIGR-HGS/Celera were able to rely on these results to construct their own map of the human genome sequence, although they did not contribute back by depositing their own results into GenBank.[119] The HGP consortium then received support from an unexpected quarterSplit sentence in two. Specifically, pharmaceutical companies such as Merck pledged to contribute their own sequence data to GenBank so as to build a public domain of sequence information.[120]

The reasons for supporting the public sequencing of the human genome were entirely pragmatic. By driving the development and growth of the human genome sequence as a public domain record, these pharmaceutical companies ensured that they were able to conduct drug discovery freely. Any research or clinical trials would not require these companies to pay

---

[114] Christopher Anderson, 'US patent application stirs up gene hunters' (1991) 353(6344) *Nature* 485–486; Rebecca S. Eisenberg, 'Genes, Patents, and Product Development' (1992) 257(5072) *Science* 903–908; Bernadine Healy, 'On Gene Patenting' (1992) 327(9) *New England Journal of Medicine* 664–668; L. Roberts, 'NIH gene patents, round two' (1992) 255(5047) *Science* 912–913.

[115] Robert M. Cook-Deegan, above n97, 311.

[116] Robert M. Cook-Deegan, above n97 311.

[117] Hallam Stevens, 'The Politics of Sequence: Data Sharing and the Open Source Software Movement' (2015) 50(4) *Information & Culture: A Journal of History* 465–503 483.

[118] Hallam Stevens, above n117, 488.

[119] Leslie Roberts, 'Controversial From the Start' (2001) 291(5507) *Science* 1182–1188 1188; Robert Cook-Deegan, Rachel A. Ankeny and Kathryn Maxson Jones, above n107, 2-3.

[120] Jorge L. Contreras and A. Jamie Cuticchia, *Bioinformatics Law: Legal Issues for Computational Biology in the Post-genome Era* (American Bar Association, 2013) 7–8; Hallam Stevens, above n117, 484-5.

licence fees or being impeded by a thicket of overlapping patents.[121] Venter responded in 1998 by promising to release a shotgun sequenced version of the human genome sequence available for free on Celera's web servers. Eventually, both the publicly funded HGP and Celera brokered a truce to publish their final results simultaneously, although even this truce eventually broke down and the two groups published in separate journals.[122] Despite both the HGP and Venter describing the controversy over the publishing order of their respective genome sequences as 'irrelevant', the dispute still had a significant impact on the way that other genomic projects were structured.

Soon after the complete human genome sequences were published, the attentions of the molecular biology community turned towards applying the human genome map in personalised medicine.[123] The International HapMap project was founded with the goal of identifying common patterns of sequence variation within the human genome.[124] In other words, the goal of this project was to identify the blocks of DNA, or haplotypes, responsible for *de novo* mutations. These haplotypes can be used to explain the differences between distinct human populations.[125] The publicly funded SNP Consortium funded the International HapMap Project and developed a public licensing system to protect HapMap data from 'parasitic patenting'.[126] This licensing system was based on previous 'free software' style licences and required all users of the International HapMap project data to verify they would not place additional limitations on the use of that data.[127] The next section will discuss the technological platforms that have evolved subsequent to the HGP and the International HapMap Project for analysing data.

---

[121] Hallam Stevens, above n117, 484.

[122] Eric S. Lander et al., 'Initial Sequencing and Analysis of the Human Genome' (2001) 409(6822) *Nature* 860–921; J. Craig Venter et al., 'The Sequence of the Human Genome' (2001) 291(5507) *Science* 1304–1351; Rebecca S. Eisenberg and Richard R. Nelson, 'Public vs. Proprietary Science: A Fruitful Tension?' (2002) 77(12, Part 2) *Academic Medicine* 1392–1399.

[123] 'Integrating Ethics and Science in the International HapMap Project' , (2004) 5(6) *Nature reviews. Genetics* 467–475; National Research Council, above n11, 34.

[124] International HapMap Consortium, 'The International HapMap Project' (2003) 426(6968) *Nature* 789–796.

[125] Claire T. Driscoll, 'NIH Data and Resource Sharing, Data Release and Intellectual Property Policies for Genomics Community Resource Projects' (2005) 15(1) *Expert Opinion on Therapeutic Patents* 1–8.

[126] Donna M. Gitter, 'Resolving the Open Source Paradox in Biotechnology: A Proposal for a Revised open Source Policy for Publicly Funded Genomic Databases' (2006) 43(5) *Houston Law Review* 1475 1509.

[127] Dianne Nicol, 'Cooperative Intellectual Property in Biotechnology' (2007) 4(1) *SCRIPT-ed* 136–151 148-9.

1.4.5 *After the HGP and HapMap - Next Generation Sequencing and The Expansion of Bioinformatics Algorithms and Databases*

Both the HGP and the International HapMap projects kickstarted the development of techniques allowing for higher sequencing throughput.[128] The need for higher sequencing throughput was driven by the intense competition between the privately and publicly funded human genome sequencing initiatives to release sequence data. This higher sequencing throughput is in turn driven by the increased computational power of sequencing hardware.[129] This trend prompted what is known as the development of 'Next Generation Sequencing (NGS)' technologies. Although there are multiple (NGS) platforms, they all differ from sequencing (which has often been described as a first generation sequencing technology) in three key ways. Firstly, NGS technologies do not rely on chemical methods of replicating DNA in bacterial cells (rDNA technology) but instead operate in a cell free system. Secondly, NGS technologies are more efficient in that they have a significantly higher throughput than standard Sanger sequencing technologies. Thirdly, the different sequences are read directly without the need for the additional step of electrophoresis or translation onto an autoradiograph.[130] This technical advance again increases the speed with which sequencing can be performed.

All NGS platforms sequence millions of small fragments of DNA in parallel by using cyclic sequencing techniques to repeatedly sequence each fragment.[131] The most prominent platform is Illumina (formerly Solexa) sequencing, named after the parent company (which has incidentally taken a very proprietary approach to establish market dominance).[132] As for the first automated sequencer platforms developed at Caltech, the inspiration for developing next generation sequencing technology is largely commercial. However, as Stevens notes, the decrease in the cost of sequencing that has followed the introduction of NGS technology has also made NGS technology more accessible for smaller labs.[133] Crucially, the surge in genomic data flowing from next generation sequencing has prompted a corresponding surge in the number of software packages and databases developed for molecular biology work. Table

---

[128] Wilhelm J. Ansorge, 'Next-generation DNA sequencing techniques' (2009) 25(4) *New biotechnology* 195–203 195-196.

[129] Elaine R. Mardis, 'The Impact of Next-Generation Sequencing Technology on Genetics' (2008) 24(3) *Trends in Genetics* 133–141 133; Mihai Pop and Steven L. Salzberg, 'Bioinformatics Challenges of New Sequencing Technology' (2008) 24(3) *Trends in Genetics* 142–149 142; Charles Lawson and Michelle Rourke, above n106.

[130] Erwin L. van Dijk et al., 'Ten Years of Next-Generation Sequencing Technology' (2014) 30(9) *Trends in Genetics* 418–426 419-420.

[131] Jay Shendure and Hanlee Ji, 'Next-Generation DNA Sequencing' (2008) 26(10) *Nature Biotechnology* 1135–1145 1135-1136.

[132] Wilhelm J. Ansorge, above n128, 196.

[133] Hallam Stevens, above n89, 206-9.

1.1 below (adapted from Contreras and Cuticcha) lists four of the most prominent open access databases.

NCBI	The NCBI databases provide open access DNA sequence and associated data (http://www.ncbi.nlm.nih.gov).
OMIM	The Online Mendelian Inheritance in Man (OMIM) database is another NCBI funded database that provides access to mendelian inheritance information regarding human genes and phenotypes.
UniProt	UniProt is a online database run by EMBL and the Swiss Institute of Bioinformatics (SIB) which provides data on protein sequences and their respective function.
PDB	The Protein Databank (PDB) is a databank run by a worldwide consortium (the World Wide Protein Databank, which has subsidiaries in Japan, the European Union and the United States) that provides a repository of three dimensional protein structures.

Table 1.1: A comparison of each of the most frequently used bioinformatics databases. Adapted from Jorge L. Contreras and A. Jamie Cuticchia, *Bioinformatics Law: Legal Issues for Computational Biology in the Post-genome Era* (American Bar Association, 2013) 12.

All of the databases in Table 1.1 contain data available for free because, like the HGP and International HapMap projects, they benefit from significant public funding.[134] However, each of these different data collections have different rules for commercial and academic users of data. In particular, EMBL has chosen to delineate between primary and secondary molecular biology databases for the purposes of determining intellectual property rules.[135] Primary databases contain data directly sourced from a particular experimental group rather than an amalgamation of data from separate sources. In contrast, secondary databases may draw on both primary and other secondary databases to draw inferences based on a combined dataset. As a result of the cost of annotation in terms of time and financial resources, secondary databases are usually subject to more strict limitations than primary databases on how downstream data is used.[136]

The proliferation of bioinformatics databases into molecular biology research as research repositories has also seen the simultaneous spread of bioinformatics into molecular biology as an interpretational tool. Software development in bioinformatics focused initially on more efficient algorithms to sequence protein sequences. However, it now encompasses all aspects

[134] Jorge L. Contreras and A. Jamie Cuticchia, above n120, 8-9; Peter W. Rose et al., 'The RCSB Protein Data Bank: Integrative View of Protein, Gene and 3D Structural Information' (2017) 45(D1) *Nucleic Acids Research* D271–D281 280.

[135] European Molecular Biology Laboratory-European Bioinformatics Institute, *Primary and Secondary Databases* (2016) *EMBL-EBI Train Online* https://www.ebi.ac.uk/training/online/course/bioinformatics-terrified/what-database/relationship-databases/primary-and-secondary-databases

[136] Jorge L. Contreras and A. Jamie Cuticchia, above n120, 10-11; Jerome H. Reichman, Tom Dedeurwaerdere and Paul F. Uhlir, above n11, 362.

of software engineering, from the development of visualisation software to networking technologies to aid with the transfer of data.[137] The former is closely related to the emerging field of systems biology, discussed in Section 1.4.4. By contrast, the latter represents an ongoing challenge within bioinformatics; that is, handling vast quantities of protein and DNA sequence data.[138] Some software is designed purely to assist with the collaborative development of software, such as the content management software Github. Platforms such as Github allow multiple authors to work on an individual project whilst tracking the various revisions that have been made to that project.[139]

In addition, the rate of genomic data production has led to increased emphasis on cloud computing as a means to meet the needs of NGS data analysis.[140] 'Cloud computing' is a generic term to describe using networked or remote servers to host software rather than relying on local computing hardware.[141] Within computational biology, cloud computing usually involves uploading data to a central server. Software installed on this server can then be used to analyse this data, as illustrated in Figure 1.2.

Whilst cloud computing may represent a solution to the 'data deluge' flowing from NGS technologies, it also carries a number of technical challenges. These challenges include the time delay taken to upload and analyse data, data loss, server downtime and data security.[142] Even the process of transferring data from the sequencing centre is a non-trivial task and may require the use of a physical hard drive.[143] There is also uncertainty as to intellectual property protection associated with data and software hosted in a cloud computing environment.[144] In particular, the issues around ownership of data in a cloud computing environment are elaborated in subsequent chapters of this thesis.

1.4.6 *Beyond Bioinformatics Databases and Software - Systems and Synthetic Biology*

As discussed in previous sections, bioinformatics algorithms and databases were developed to introduce quantitative measures for understanding the operation of protein and (later) DNA

---

[137] National Research Council, above n11, 52-53.

[138] Jerome H. Reichman, Tom Dedeurwaerdere and Paul F. Uhlir, above n11, 362-363.

[139] See <https://github.com/>

[140] Lincoln D. Stein, 'The Case for Cloud Computing in Genome Informatics' (2010) 11(5) *Genome Biology* 207.

[141] Michael Armbrust et al., 'A View of Cloud Computing' (2010) 53(4) *Communications of the ACM* 50–58.

[142] Arnon Rosenthal et al., 'Cloud Computing: A New Business Paradigm for Biomedical Information Sharing' (2010) 43(2) *Journal of Biomedical Informatics* 342–353 345-346.

[143] Krithika Bhuvaneshwar et al., 'A Case Study for Cloud Based High Throughput Analysis of NGS Data Using the Globus Genomics System' (2015) 13(Supplement C) *Computational and Structural Biotechnology Journal* 64–74 67.

[144] Jerome H. Reichman, Tom Dedeurwaerdere and Paul F. Uhlir, above n11 362-363.

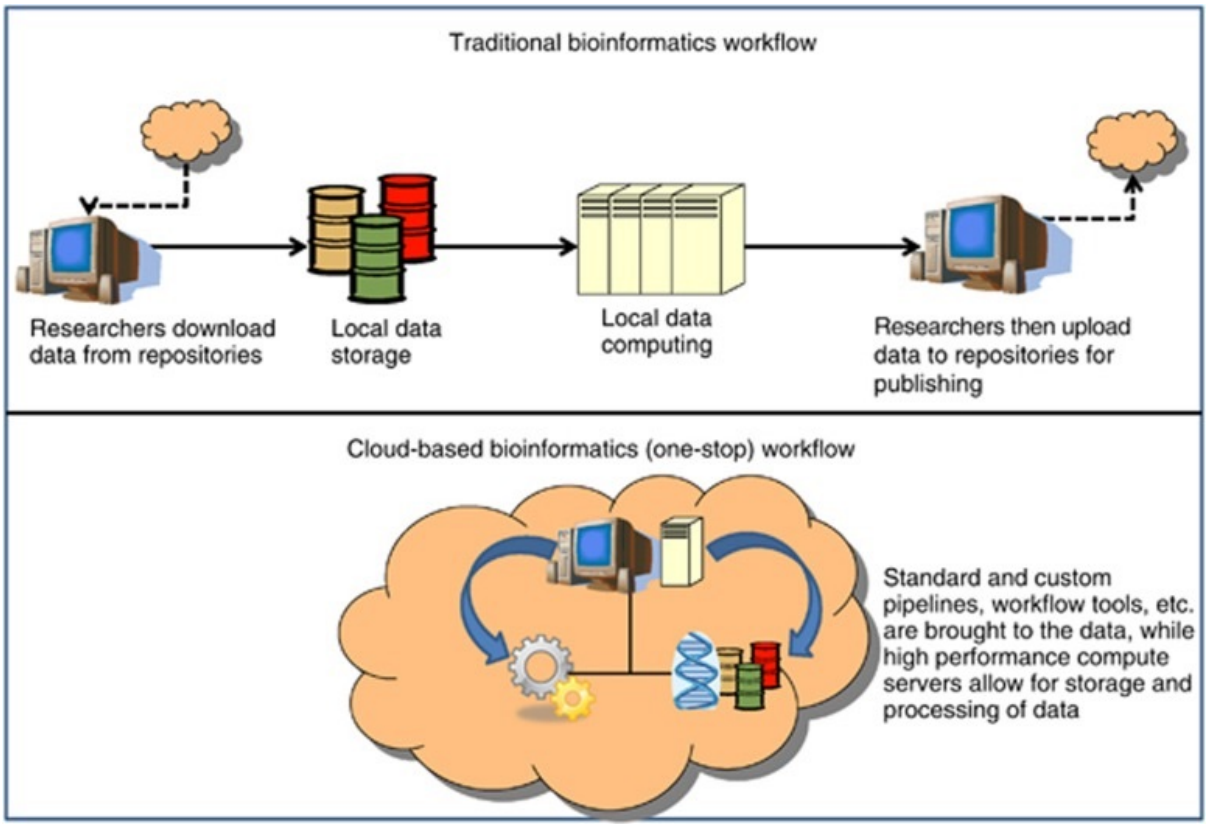

Figure 1.2: A contrast between the traditional bioinformatics workflow and a cloud based workflow, where all steps of the workflow are integrated onto the one high performance machine (From Edward S. Dove, Yann Joly, Anne-Marie Tassé, Bartha M. Knoppers Public Population Project in Genomics and Society (P3G) International Steering Committee, International Cancer Genome Consortium (ICGC) Ethics and Policy Committee, 'Genomic cloud computing: legal and ethical points to consider' (2015) 23(10) *European Journal of Human Genetics*, 1271, 1272. Reproduced with permission.)

sequences. However, these algorithms focused on understanding the function of these particular proteins in isolation. Projects such as the HGP and the HapMap Project prompted molecular biologists to consider how proteins and DNA sequences are responsible for the functioning of whole biological systems.[145] In biomedicine, these systems oriented approaches are grounded in the notion that disease affected proteins and gene regulatory networks operate differently to their unaffected counterparts.[146] To this end, systems biology software focuses on modelling biological systems as a whole, and builds on existing computational biology tools to understand cellular processes such as gene expression and metabolism.[147] Accordingly, the emphasis in systems biology rests on attempting to predict the functioning of biological systems, rather than attempting to merely focus on why biological systems work a particular way. This shift away from reductionism towards

[145] Pascal Kahlem and Ewan Birney, 'Dry Work in a Wet World: Computation in Systems Biology' (2006) 2(1) *Molecular Systems Biology* 40.

[146] Leroy Hood et al., 'Systems Biology and New Technologies Enable Predictive and Preventative Medicine' (2004) 306(5696) *Science* 640–643 460.

[147] National Research Council, above n11.

attempting to holistically model biological systems also endeavours to incorporate the impact of environmental factors.[148] Such a holistic approach is highly dependent on both open access to data to help model a biological system and software to provide the necessary infrastructure to integrate this data together.[149]

Closely related to the idea of modelling a biological system using software and data is the notion of engineering emergent life forms. Although the idea of using rDNA is not new, synthetic biology is directed towards the goal of developing modular components which are standardised and interchangeable.[150] This engineering approach to biology attempts to replicate object-oriented approaches to programming, where modular components can be interchanged in a standardised fashion to build larger software programs. As with systems biology, synthetic biology is founded on computational biology software that can be used to identify and visualise genome function.[151] This engineering approach to biology has the potential to realise many of the social benefits of computational biology in a diverse range of fields as described in Section 1.2 of this Chapter.[152] In particular, the NRC's report notes that highly optimised cells or organisms may have enormous benefits in environmental science, biomedicine and energy production.[153] However, because of the inherent complexity in biological systems, the challenges of engineering biology are significantly greater than merely modelling biological systems with software.[154] These challenges suggest that despite an inclination amongst the emergent synthetic biology community towards open access models to encourage greater standardisation, some other form of reward system is required beyond the open access model that has dominated computational biology so far.[155]

---

[148] Lee Hood, 'A Personal Journey of Discovery: Developing Technology and Changing Biology' (2008) 1(1) *Annual Review of Analytical Chemistry* 1–43 21-22.

[149] Jane Calvert and Joan H. Fujimura, 'Calculating life? Duelling Discourses in Interdisciplinary Systems Biology' (2011) 42(2) *Studies in History and Philosophy of Biological and Biomedical Sciences* 155–163.

[150] Jane Calvert, 'The Commodification of Emergence: Systems Biology, Synthetic Biology and Intellectual Property' (2008) 3(04) *BioSocieties* 383–398 384; Maarten Boudry and Massimo Pigliucci, above n71, 661.

[151] Linda J. Kahl and Drew Endy, 'A survey of enabling technologies in synthetic biology' (2013) 7(1) *Journal of Biological Engineering* 13.

[152] Drew Endy, 'Foundations for Engineering Biology' (2005) 438(7067) *Nature* 449–453 449.

[153] National Research Council, above n11, 62-63.

[154] Drew Endy, above n152, 449-450; Maarten Boudry and Massimo Pigliucci, above n71 661.

[155] Joachim Henkel and Stephen M. Maurer, 'The Economics of Synthetic Biology' (2007) 3(1) *Molecular Systems Biology* 117 2.

1.5 PRELIMINARY CONCLUSIONS ON THE NATURE OF BIOINFORMATICS AND IMPLICATIONS FOR THE PRESENT STUDY

Sections 1.2, 1.3 and 1.4 have described in detail the evolution of computational biology and the sociological influences from both molecular biology and computer science. In particular, this historical account has demonstrated how computational biologists have prioritised the importance of open access (or open source, to use the computer science nomenclature) approaches to computational biology. The next section will now discuss the two predominant competing considerations in the development of open access computational biology systems; reproducibility and sustainability.

1.5.1 *Reproducibility and Computational Biology*

As discussed in Section 1.2, the original justification for introducing quantitative methodology into evolutionary (and later molecular) biology was to inject objective measures of comparison into what had traditionally been a subjective field.[156] As the power of computational techniques exponentially increased, so too did the need to ensure that these results are reproducible. In other words, there is an increasing emphasis on computationally driven research to explain how to achieve the results reported.[157] The need for reproducibility has increased alongside the increase in computational power and the amount of data (including genomic data) that can be processed with this software.[158]

Although there is significant literature on what statistical measures are appropriate for determining whether research is 'reproducible' (which is beyond the scope of this thesis), the need for reproducible results may also have significant economic consequences. As this thesis will elaborate in subsequent chapters, these consequences include the economic cost of irreproducible research. The economic cost of irreproducible research is particularly significant for applied biomedical research such as drug discovery.[159] In addition, there may be follow on costs to patent holders who have sought patents based on irreproducible research.[160] There are also social concerns such as invasion of privacy flowing from data mining of large sets of data, particularly with individual data such as private health records.[161]

---

[156] Joel Hagen, above n28, 359.

[157] Vincent J. Carey and Victoria Stodden, 'Reproducible Research Concepts and Tools for Cancer Bioinformatics' in Michael F. Ochs, John T. Casagrande and Ramana V. Davuluri (eds.), *Biomedical Informatics for Cancer Research* (Springer US 2010)149–175 165-7.

[158] Victoria Stodden et al., above n87, 1240.

[159] Leonard P. Freedman, Iain M. Cockburn and Timothy S. Simcoe, 'The Economics of Reproducibility in Preclinical Research' (2015) 13(6) *PLOS Biol* e1002165 3.

[160] Jacob S. Sherkow, 'Patent Law's Reproducibility Paradox' (2016) 66(4) *Duke Law Journal* 875-82.

[161] Victoria Stodden, 'Enabling Reproducibility in Big Data Research: Balancing Confidentiality and Scientific

These important social and economic implications raise the question of how best to achieve reproducibility in computational biology from both a technical and legal perspective. In particular, the concept of reproducibility, which has gained significant currency in the computational biology community, involves releasing both the software and data associated with publications for inspection.[162] However, there is debate within the scientific community as to whether complete open access to software necessarily ensures reproducibility or replicability.[163]

To this end, critics distinguish between *replicability* (or merely achieving the same results through repeating the engineering process) and *reproducibility* (actively using different methods to achieve similar outcomes).[164] In addition, the question of open source licensing to ensure reproducibility raises other intellectual property issues. In particular, as discussed previously, the vast majority of computational biology software is developed to support proprietary sequence analysis platforms.[165] The divide between open source and proprietary development platforms is relevant to another competing consideration in computational biology identified in this chapter, that of sustainability.

1.5.2 *Sustainability and Bioinformatics*

When this thesis refers to the concept of sustainability, it refers to the means by which computational biology software and infrastructure is maintained.[166] As discussed previously, early computational biology databases, as well as large scale computational biology projects like the HGP and the International HapMap Project, were all largely supported through public funding regimes.[167] Sections 1.4.2 also discussed the reputation based 'gift economies' in experimental biology research. The nature of these sharing economies undermined attempts by Dayhoff and her collaborators to establish a reimbursement system for the Atlas. As subsequent chapters will explore, similar arguments have also been raised regarding the economic motivations for programmers to volunteer their time for open source software

---

Transparency' in *Privacy, Big Data and the Public Good* (Cambridge University Press 2014)112 113.

[162] Sören Sonnenburg et al., 'The Need for Open Source Software in Machine Learning' (2007) 8(Oct) *Journal of Machine Learning Research* 2443–2466 2456; Victoria Stodden, 'Open Science: Policy Implications for the Evolving Phenomenon of User-Led Scientific Innovation' (2010) 9(1) *Journal of Science Communication* 1–8 4.

[163] Hans E. Plesser, 'Reproducibility vs. Replicability: A Brief History of a Confused Terminology' (2018) *Frontiers in Neuroinformatics* 2.

[164] Joanna Lewis et al., 'Where Next for the Reproducibility Agenda in Computational Biology?' (2016) 10(1) *BMC Systems Biology* 52 3.

[165] Wilhelm J. Ansorge, above n128, 196-99.

[166] Amrita Mishra, Paul N. Schofield and Tania M. Bubela, above n5, 292-3.

[167] Robert Cook-Deegan, Rachel A. Ankeny and Kathryn Maxson Jones, above n107, 6.3.

projects.[168]

However, where research has an applied or translational impact, the need for ongoing funding often necessitates commercialisation of an invention beyond an academic environment. This effect can be seen through the early development of sequencing platforms and techniques at Caltech, as well as the proliferation of proprietary NGS platforms.[169] Moreover, even open source systems require ongoing maintenance and support, which can rapidly decline without adequate incentives to continue to contribute to a project. This question of sustainability is an equally important consideration for publicly funded as well as privately funded research.

By way of example, McLeod notes that in 2015, Australian government and non for profit researchers spent 407 million Australian dollars, or 40 percent, of their total research and development expenditure, on Information Communications and Technology (ICT) research. Likewise, Australian private sector firms in 2015 also spent 7.7 billion Australian dollars, or 41 percent of their total research and development expenditure on ICT research. Despite this enormous outlay, however, ICT and electrical engineering has the highest cost per publication of all science, technology, engineering and management (STEM) fields.[170] Moreover, within the context of Australian and New Zealand bioinformatics there are concerns about the fragmented nature of the research field. These concerns were reflected in the 2009 National Bioinformatics Strategy, which called for a dedicated maintenance strategy for bioinformatics.[171]

Perhaps the most frequently relied upon means to guarantee sustainability is to seek intellectual property protection for computational biology software, methods and data. For example, Dayhoff's attempts to seek reimbursement for the publication of the Atlas were founded in copyright law. Likewise, Gregory Kirsch and Charley Brown note that many of the developers of automated sequencing platforms, including Illumina and Oxford Nanopores, now seek patent protection for the software bundled with their sequencing platforms.[172] In

---

[168] Martin Michlmayr, above n3, 23; Robert R. Downs et al., 'Community Recommendations for Sustainable Scientific Software' (2015) 3(1) *Journal of Open Research Software* 3.

[169] Wilhelm J. Ansorge, above n128 196-99.

[170] Annette McLeod, *Returns on Investment: Considerations on Publicly Funded ICT Research and Impact Assessment* (PhD thesis, University of Melbourne, 2016) 7, 18 <http://minerva-access.unimelb.edu.au/handle/11343/124272>.

[171] Biotechnology Australia, *Australian Bioinformatics Network Final Report* (2009) 6-8 <http://australianbioinformatics.net/storage/downloads/Australian%20Bioinformatics%20Network%20Project%20-%20Final%20Report%20-%2020080118.pdf>.

[172] Gregory J. Kirsch and Charlie F. Brown, 'Software patents' in Jorge L. Contreras and A. Jamie Cuticchia (eds.), *Bioinformatics Law: Legal Issues for Computational Biology in the Post-genome Era* (American Bar Association 2013) 56-9.

addition to these traditional uses of intellectual property, the open source approaches that evolved from open sharing regimes from academic software rely on inversions of intellectual property protection. Specifically, these strategies include open source licensing and community norms to protect openly licensed software from being appropriated.[173] However, all licensing mechanisms, irrespective of whether they are proprietary or open, must operate within the confines of patent and copyright laws. Although there are international treaties to govern the domestic implementation of copyright and patent law, there are still considerable differences in the national operation of these laws, particularly with respect to software and data.

1.6 CONCLUSION

This chapter has charted the evolution of bioinformatics as an emergent scientific discipline by focussing on two main paradigm shifts in scientific research. The first of these shifts has been the transition in molecular biology research towards a quantitative discipline reliant on numerical techniques to conduct comparative evolutionary and molecular biology research. In particular, the discovery of protein and genome sequences prompted the use of both statistical methods to analyse biological phenomena as well as the development of specific platforms for reading protein and DNA sequences. The shift towards quantitative models of computational biology has been matched by the corresponding emergence of computer science as a scientific discipline in its own right. In particular, the evolution of software engineering, operating systems and networking technologies have driven the increased use of specialised software in laboratory research and the institutionalisation of scientific software developers to produce that software. Bioinformatics researchers are now embedded in mainstream molecular biology research, either as internal support staff or as external collaborators on cross-institutional projects, as demonstrated by large scale computational biology projects such as the HGP and the International HapMap Project.

In addition, the fact that the free and open source software movement evolved from scientific research communities has led to a long history of both open source development practices being embedded in bioinformatics research and a resistance against the introduction of strong proprietary protection for software used in academic bioinformatics research. Accordingly, from a technical perspective there are two conflicting ways in which copyright and patent laws can either positively or negatively impact the development of bioinformatics software. Firstly, copyright and patent laws (as well as copyright and patent licensing) may influence the degree to which bioinformatics software is reproducible; that is, the degree to which the results generated using bioinformatics software can be recreated. Secondly,

[173] Siobhán O'Mahony, 'Guarding the Commons: How Community Managed Software Projects Protect their Work' (2003) 32(7) *Research Policy* 1179–1198 1183.

competing against the need for reproducibility is the requirement of sustainability; that is, whether a bioinformatics project can attract ongoing development support. Chapters Two and Three will now turn to address the boundaries of each of these laws on a comparative jurisdictional basis.

Chapter 2

COPYRIGHT IN CODE: A COMPARATIVE JURISDICTIONAL ANALYSIS OF DIVERGENT COPYRIGHT LAWS ON SOFTWARE

2.1 INTRODUCTION

This chapter frames the concepts around the relationship between copyright law and computational science. A key criticism of copyright law has been a failure to adapt and evolve in the face of new and emerging technologies. In particular, software has created consternation for both domestic legislators and international negotiators regarding the scope of protection for copyright (or indeed whether copyright protection should extend to software at all).[1] Closely related to the extent of protection for software is the question of protection for data and whether copyright protection should extend to facts and collections of data, which (as discussed in Chapter One) can now be easily compiled using software.

The chief reason for this uncertainty is that software and data invoke the tension which lies at the heart of copyright law. Specifically, there is friction between protecting the interests of creators by providing an avenue to prevent the unauthorised use of their creative works, and maintaining of a viable public domain with which to encourage further creativity. These tensions are expressed through exceptions to the rights held by copyright owners, including fair use (or fair dealing) exceptions and reasonable limitations on what can be copyrighted (such as compilations of data and software interfaces).[2] Due to the flexibility inherent in international law with respect to how signatories may implement copyright laws, there are significant differences between jurisdictions in how each of these exceptions operate. For the purposes of this thesis, understanding how these divergent copyright laws work is important for understanding how the alternative copyright framework of open source software can co-exist alongside more traditional models of copyright protection.

The aim of this chapter is therefore to chart the boundaries of copyright protection, as well as the economic justifications for different forms of copyright protection. This chapter is therefore split into four sections. Section 2.2 examines the boundaries of copyright law, as well as accounting for how software became copyrightable under domestic and international law. Section 2.3 considers the scope of copyright protection for software under the US copyright statute, and how this differs from copyright protection in Europe, Australia and New Zealand. Section 2.3 also examines the separate *sui generis* right for databases and compilations of facts that exists at the supranational level in the EU, as well as how this *sui generis* regime compares to copyright protection for data in the US, Australia and New Zealand. Section 2.4 addresses the legal distinction between 'closed' and 'open' source licences that rests at the

---

[1] James Bessen and Michael J. Meurer, *Patent Failure: How Judges, Bureaucrats, and Lawyers Put Innovators at Risk* (Princeton University Press, 2009) 187.

[2] Sally Weston, 'Software Interfaces - Stuck in the Middle: The Relationship Between the Law and Software Interfaces in Regulating and Encouraging Interoperability.' (2012) 43(4) *IIC International Review of Intellectual Property and Competition Law* 427–450 428; Isabella Alexander, 'Manacles upon Science: Re-Evaluating Copyright in Informational Works in Light of 18th Century Case Law' (2014) 38(2) *Melbourne University Law Review* 317–361 355.

heart of the research in this thesis and which first emerged in the US, but later spread to other jurisdictions. In particular, there has been substantial judicial commentary on the enforceability of open source licences in the US and the EU. However, there has been comparatively little judicial commentary on the enforcement of open source licences in Australia and New Zealand. Finally, Section 2.5 targets the scope of fair use exemptions under copyright law. It specifically compares the relatively broad protection provided by the fair use doctrine in the US and the narrow fair dealing exemptions provided under EU, Australia and New Zealand copyright law.

2.2 THE BOUNDARIES OF COPYRIGHT LAW

2.2.1 *A Brief History and Definitional Grounding in Copyright Law*

Copyright law is designed to protect tangible works which are produced by one or more creators, and fixed in a tangible medium, as opposed to ideas which are purely functional in nature, or are not tied to any fixed medium.[3] The 'idea-expression' divide is particularly important in considering the extent of copyright protection for software.[4] Copyright gives an author the exclusive right to reproduce the copyrighted work, to produce adaptations of the original work or to transfer these rights to another creator to perform the same actions.[5] The existence of copyright law is based on an assumption that by offering an incentive to produce expressions, artists will continue to do so in a way that is socially beneficial.[6] The term 'socially beneficial' in turn raises a number of questions about the scope, length and limits of copyright protection, which in turn are influenced by international laws.

Although copyright laws are implemented on a domestic level, there are a number of international agreements which define minimum standards for copyright protection. These specifically include the *Berne Convention* and the *TRIPS Agreement*.[7] These treaties mandate that signatories establish penalties for the infringement of the exclusive rights available under copyright. This infringement can occur through either direct purposeful infringement or

---

[3] Suzanne Scotchmer, *Innovation and Incentives* (MIT Press, 2004) 115.

[4] Noam Shemtov, 'Circumventing the Idea/Expression Dichotomy: The Use of Copyright, Technology and Contract to Deny Access to Ideas' in Guido Westkamp (ed.), *Emerging Issues in Intellectual Property: Trade, Technology and Market Freedom : Essays in Honour of Herchel Smith* (Edward Elgar Publishing 2007) 89.

[5] Reto M. Hilty and Sylvie Nérisson, *Balancing Copyright - A Survey of National Approaches* (Springer Science & Business Media, 2012) 5.

[6] Richard Watt, 'Copyright and Contract Law: Economic Theory of Copyright Contracts The Relationship between Copyright and Contract Law' (2010) 18(1) *Journal of Intellectual Property Law* 173–206; Richard Watt, *Handbook on the Economics of Copyright: A Guide for Students and Teachers* (Edward Elgar Publishing, 2014) 10.

[7] *Berne Convention for the Protection of Literary and Artistic Works*, opened for signature 9th September 1886, 1161 UNTS 30 (entered into force 4th December 1887).

secondary infringement by aiding and abetting infringement.[8] Finally, they provide a three step test for determining the scope of copyright exceptions and limitations in special cases. These special exceptions apply where the use does not conflict with normal exploitation of the work and does not limit the rights of the legitimate rights holder.[9] This question of how different jurisdictions handle copyright exceptions is something that Parts 2.3 to 2.5 of this chapter will consider in greater detail.

One of the most crucial aspects of these treaties is their role in setting minimum standards for the subject matter that can be protected by copyright. In particular, Article 9(2) of the *TRIPS Agreement* codifies existing common law principles that copyright only extends to expressions and not 'ideas, procedures, methods of operation or mathematical concepts as such'. Article 10.1 then goes on to state that computer programs, 'whether in source or object code', shall be protected as literary works under the 1971 redraft of the Berne Convention. But how did international law reach this point? The next section traces how copyright protection became the default mechanism for protection of software source code.

2.2.2 *How Software Became Copyrightable*

Pamela Samuelson notes that arguments about copyright protection for software first began in the 1960s and 1970s. During this time, lawyers and economists debated over what form (if any) of legal protection was or should be available for computer programs.[10] At the time the US software industry was flourishing without the support of copyright protection and computer programs were seen as having no independent significance from the hardware they were stored on.[11] As such, lawmakers deemed it unnecessary to extend copyright protection to software.[12] In addition, the availability of copyright protection for software was and has

---

[8] Mihály Ficsor, 'Collective Management of Copyright and Related Rights from the Viewpoint of International Norms and the Acquis Communautaire' in Daniel J. Gervais (ed.), *Collective Management of Copyright and Related Rights* (Kluwer Law International 2010) 39.

[9] *Berne Convention for the Protection of Literary and Artistic Works*, opened for signature 9th September 1886, 1161 UNTS 30 (entered into force 4th December 1887) article 5(2), article 7(1), article 9(2); *Marrakesh Agreement Establishing the World Trade Organization, annex IC, The Agreement on Trade Related Aspects of Intellectual Property Rights ('TRIPS Agreement')*, opened for signature 15th April 1994, 1867 UNTS 3 (entered into force 1st January 1995) article 12, article 13; Christophe Geiger, Daniel Gervais and Martin Senftleben, 'The Three-Step Test Revisited: How to Use the Test's Flexibility in National Copyright Law' (2013) 29(3) *American University International Law Review* 581–626.

[10] Pamela Samuelson, 'A Square Peg in a Round Hole? Copyright Protection for Computer Programs' in Brad Sherman and Wiseman (eds.), *Copyright and the Challenge of the New* (Kluwer Law International 2012) 251.

[11] Kenneth W. Dam, 'Some Economic Considerations in the Intellectual Property Protection of Software' (1995) 24(2) *Journal of Legal Studies* 321–378 327.

[12] Stephen Breyer, 'The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs' (1970) 84(2) *Harvard Law Review* 281–351 347-8; Pamela Samuelson, 'The Uneasy Case for Software Copyrights Revisited' (2010) 79(6) *George Washington Law Review* 1746 1746-1747.

continued to be problematic from a legal perspective. This problematic nature arises from the fact that copyright protection has only been available for expressed as opposed to functional works. This prohibition on protection for functionality goes to the heart of concerns around copyright protection for software: its dual literary and functional nature raises vexing issues. On the one hand, source code, which is characterised by syntax and operators, is clearly a literary work. On the other hand, other aspects of computer programs such as application programming interfaces[13] are largely functional. Therefore, these programs therefore may not amount to sufficiently original expressions to warrant copyright protection.[14]

For these reasons, the US Copyright Office was initially strongly opposed to extending copyright protection for software.[15] Yet over time it became apparent that excluding software from copyright protection was not congruent with the economic reality of software development. For example, because until 1978 International Business Machines (IBM) released all of their mainframe computer software for free. Accordingly, Fujitsu were able to reverse engineer this software and produce mainframe machines that were backwards compatible with IBM's software. This reverse engineering undermined IBM's domestic and international hardware sales.[16] To resolve this lack of protection, the US Congress convened the National Commission on New Uses of Copyrighted Works (CONTU) in 1974. This Commission was convened concurrently to similar investigations in Europe and Japan, both of which were contemplating a *sui generis* form of protection for software. This *sui generis* regime would be equivalent to other forms of *sui generis* protection, such as plant breeder's rights or circuit board protection.[17]

However, a majority of the CONTU panel ultimately concluded that the US needed to offer copyright protection for software. CONTU further concluded that courts should ultimately be left to decide which aspects of software were protected literary works or unprotected functional creations.[18] CONTU's recommendations resulted in the amendment of the US Copyright Act

---

[13] APIs, as discussed in Section 1.3.2 of Chapter One

[14] Gustavo Ghidini and Emanuela Arezzo, 'Dynamic Competition in Software Development: How Copyrights and Patents, and Their Overlapping, Impact on Derivative Innovation' (2013) 3(4) *Queen Mary Journal of Intellectual Property* 278–295 280.

[15] *Data Cash Systems, Inc. v JS &a Group, Inc.* 480 480 F. Supp. 1063 (1979); Pamela Samuelson, above n10, 252.

[16] Pamela Samuelson, 'IBM's Pragmatic Embrace of Open Source' (2006) 49(10) *Communications of the ACM* 21–25 23; *IBM v Fujitsu* No. 13T-117-0636-85 American Arbitration Ass'n Commercial Arbitration Tribunal 4 (1987); Anita Stork, 'The Use of Arbitration in Copyright Disputes: IBM v. Fujitsu' (1988) 3(2) *Berkeley Technology Law Journal* 241.

[17] Dennis S. Karjala, 'Lessons from the Computer Software Protection Debate in Japan' (1984) *Arizona State Law Journal* 53–82; Giorgio Fabio Colombo and Matteo Dragoni, 'The Legal Protection of Software in Japan—An Original Model?' in Giuseppe Bellantuono and Fabiano Teodoro Lara (eds.), *Law, Development and Innovation* (Springer 2016)67–88 71-2, 76.

[18] Gerardo Con Dıaz, 'The Text in the Machine: American Copyright Law and the Many Natures of Software, 1974–1978' (2016) 57(4) *Technology and Culture* 753–779 756, 777.

in 1980 to include a specific definition of a computer program as:

> a set of statements or instructions to be used directly or indirectly to bring about a certain result

This definition extended copyright protection to software (including both object and source code). The amendment also included specific exemptions for copying, making backup copies of, and reselling computer programs.[19] In addition to extending copyright protection to software, an amendment flowing from CONTU was to formalise the boundaries of the fair use doctrine with respect to software copyright. The doctrine of fair use was originally imported from the United Kingdom (UK) as a common law doctrine.[20] However, it was eventually codified in the US via the *Copyright Act* of 1976.[21] This statutory doctrine includes a four factor test for weighing whether allegedly infringing use amounts to fair use:

1. the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
2. the nature of the copyrighted work;
3. the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
4. the effect of the use on the potential market for or value of the copyrighted work[22]

Initially, the market harm factor was considered the predominant factor in determining whether there had been copyright infringement; in other words, whether there had been a commercial or non commercial use of the copyrighted materials.[23] Section 2.3.4 below discusses how this test has evolved over time and how the transformative use factor has become the predominant factor in determining whether there has been fair use.[24] Further, with respect to copyright infringement and software (and not only the question of infringement of

---

[19] *Copyright Act 1976* (US) § 117; *Computer Software Copyright Act of 1980 Pub. L. No. 96-517, 94 Stat. 3015 1980*.

[20] Gideon Parchomovsky and Philip J. Weiser, 'Beyond Fair Use' (2010) 96(2) *Cornell Law Review* 91–138 490gyleswilcox1740[99.

[21] *Copyright Act 1976* (US) § 107; Matthew Sag, 'The Prehistory of Fair Use' (2010) 76(4) *Brooklyn Law Review* 1371–1412 1411.

[22] *Copyright Act 1976* (US) § 107.

[23] Pierre N. Leval, 'Toward a Fair Use Standard' (1989) 103(5) *Harvard Law Review* 1105–1136 1111; Jane C. Ginsburg, 'Conflicts of Copyright Ownership between Authors and Owners of Orignial Artworks: An Essay in Comparative and International Private Law' (1992) 17(4) *Columbia-VLA Journal of Law & the Arts* 395–416 401.

[24] *Campbell v Acuff-Rose Music* 510 U.S. 569 (1994), 569; *Cambridge University Press v Patton* 769 F.3d 1232 (11th Cir. 2014), 1262.

software copyright but copyright infringement *using* software), CONTU was left with a troubling determination as to how fair use should be best moulded to fit within these boundaries. CONTU eventually reached the conclusion that computer programs should be subject to the same rules on fair use and systematic copying as any other creative work. However, the Commission's pronouncements have been contradicted by courts, which will be discussed in further detail below.[25] For now, the most significant pronouncement from CONTU (as far as software copyright was concerned) was the extension of copyright protection to software.

2.2.3 *Diffusion of Copyright Protection for Software*

Due to the economic pressure exerted by the US as a major trading partner for Japan, the latter abandoned its plans for a *sui generis* regime for protecting software and introduced copyright protection for software.[26] This development was followed by software being formalised as copyrightable subject matter in the TRIPS Agreement.[27] As a result, signatories (including Australia and New Zealand) were compelled to introduce copyright protection for software (discussed in further detail in Section 2.3.3).[28] Once copyright became the formal intellectual property right for software, there was a massive surge in the number of software developers registering copyright in software programs. This trend was particularly observable in the US, where the Copyright Office was accepting over five thousand yearly registrations by the mid-1980s.[29]

The political influence that the US (as the home of many of the world's leading software development companies and computer science universities) had in driving and formalising software copyright protection was twofold. First, foreign investors were driven to seek copyright protection in the US if they wished to release their software in the US and keep pace with domestic developers.[30] Secondly, to prevent the leakage of software engineering

---

[25] Arthur R. Miller, 'Copyright Protection for Computer Programs, Databases, and Computer-Generated Works: Is Anything New Since CONTU' (1992) 106(5) *Harvard Law Review* 977–1073 1022.

[26] (Ministry of International Trade and Industry, Industrial Structure Council Information Industry Committee, *On the Appropriate Basic Consolidation for Software - Toward the Establishment of Legal Protection of Software*, Interim Reporrt (1983); Pamela Samuelson, above n10, 258)

[27] *Marrakesh Agreement Establishing the World Trade Organization, annex IC, The Agreement on Trade Related Aspects of Intellectual Property Rights ('TRIPS Agreement')*, opened for signature 15th April 1994, 1867 UNTS 3 (entered into force 1st January 1995) Article 10.1.

[28] Kenneth C. Shadlen, Andrew Schrank and Marcus J. Kurtz, 'The Political Economy of Intellectual Property Protection: The Case of Software' (2005) 49(1) *International Studies Quarterly* 45–71 45.

[29] Note that under the TRIPS Agreement and the Berne Convention, signatories may or may not adopt a system of copyright registration. The US is one of the few jurisdictions that has a system of copyright registration. See (Gerardo Con Dıaz, above n18, 770)

[30] Kenneth C. Shadlen, Andrew Schrank and Marcus J. Kurtz, above n28, 52.

know-how to overseas markets where copyright protection was not available, the US exerted bilateral pressure on other nations to standardise the regulation of software copyright.[31] Consequently, copyright offices and courts began to turn to the issue of how far copyright protection extended for software. The way that courts and legislators in different jurisdictions have handled these issues will be discussed in the next section of this chapter.

2.3 THE SCOPE OF COPYRIGHT PROTECTION FOR SOFTWARE

2.3.1 *Copyright for Software in the US*

Following the introduction of copyright protection for software in the US, the main question before the courts was whether copyright protection for object code extended to both the *functional* and *literal* aspects of that object code. The reasons for this legal question are largely attributable to the changing nature of software as described in Section 1.3.2 of Chapter One. In particular, the rise of interfaces, libraries and operating systems and the gradual shift away from procedural towards object oriented programming meant that for developers, the functional nature of software became more valuable than the literal source code.[32] In addition, advances in reverse engineering made it easier to reverse engineer object code. This strategy could be used to replicate the functionality of a computer program without infringing copyright by copying source code, as a form of 'forward engineering'.[33] This shift in programming practices prompted a reconsideration of the boundaries of copyright law. This shift began in earnest in 1979, where the 7th Circuit Court of Appeals in *Data Cash Systems, Inc. v. JSA*[34] rejected extending copyright protection to object code on read only memory (ROM)[35] on the grounds that the 'structure, sequence and operation' of computer programs was purely functional and therefore fell outside the limits of copyright protection as provided by *Baker v Selden*.[36]

However, in the subsequent case of *Apple v Franklin*,[37] the 3rd Circuit Court of Appeals

---

[31] Keith Maskus, 'The New Globalisation of Intellectual Property Rights: What's New This Time?' (2014) 54(3) *Australian Economic History Review* 262–284 264.

[32] Pamela Samuelson and Robert J. Glushko, 'Comparing the Views of Lawyers and User Interface Designers on the Software Copyright Look and Feel Lawsuits' (1989) 30(1) *Jurimetrics Journal* 121–140 129.

[33] Elliot J. Chikofsky and James H. Cross, 'Reverse engineering and design recovery: a taxonomy' (1990) 7(1) *IEEE Software* 13–17 14-5.

[34] *Data Cash Systems, Inc. v JS &a Group, Inc.* 480 480 F. Supp. 1063 (1979).

[35] Read only memory refers to a type of non volatile memory which can be used for relatively permanent storage (as opposed to RAM, which is volatile and can only be used for transitory storage). See (Y. Tarui, Y. Hayashi and K. Nagai, 'Electrically Reprogrammable Nonvolatile Semiconductor Memory' (1972) 7(5) *IEEE Journal of Solid-State Circuits* 369–375)) that is read only in normal operation but can be overwritten.

[36] *Baker v. Selden* 101 U.S. 99 (1880).

[37] *Apple Computer, Inc. v Franklin Computer Corp.* 714 F. 2d 1240 (3d Cir. 1983).

held that copyright was available for both an operating system's source and object code. The Court of Appeals therefore rejected the defendant's argument that they needed to reverse engineer the plaintiff's software to produce compatible hardware. Samuelson and Mark Lemley acknowledge that the decision in *Apple v Franklin* was correct insofar that exact copying of object code amounted to copyright infringement. However, they argue that *Apple v Franklin* created a troubling precedent that broadened copyright protection to both broad functional and graphical features of software.[38] In addition, *Apple v Franklin* placed the development of interoperable software outside an established exception to copyright infringement, thereby undermining the development of modular software.[39] Emboldened by their success in *Apple v Franklin*, Apple pursued Microsoft for copying the 'look and feel' of Apple's Mac OS operating system.

In both the *Apple v Microsoft*[40] series of cases, as well as the coincidental *Computer Associates v Altai*[41] series of cases, both appeal courts rejected the extension of copyright to the 'structure, sequence and organisation of computer programs'. In *Computer Associates v Altai*, the 2nd Circuit Court of Appeals were confronted with a case where the applicant's code was misappropriated by a former employee upon being offered employment with the applicant. The respondent argued that despite the purge of the tainted code, there were still substantial similarities between the two code bases. The Court of Appeals substituted the test from *Apple v Franklin* with a three-step test for software copyright infringement. The first step requires constructing a hierarchy of abstractions for the allegedly infringed software. This process involves starting with the high-level concepts expressed in the overarching algorithm and working down to the literal aspects of the software. The second step requires using the abstractions to identify any elements of the program that do not qualify for copyright protection. These abstractions may include efficiency aspects, design choices and widely accepted programming practices that are in the public domain. The third step requires comparing the remaining elements of the software with the defendant's software to determine infringement.[42] The Court of Appeals held that on the facts, infringement by copying software functionality had not been established. The Court of Appeals further held that extending copyright to abstract elements copied would also be inconsistent with more recent Supreme

---

[38] *Whelan Associates, Inc. v Jaslow Dental Laboratory, Inc.* 797 F.2d 1222 (3d Cir. 1986); Mark A. Lemley, 'Convergence in the Law of Software Copyright' (1995) 10(1) *High Technology Law Journal* 1–34 7; Pamela Samuelson, above n10, 260-1.

[39] Pamela Samuelson, 'Should Economics Play a Role in Copyright Law and Policy' (2003) 1(1-2) *University of Ottawa Law & Technology Journal* 1–22 17-19.

[40] *Apple Computer, Inc. v Microsoft Corp.* 799 F.Supp. 1006 (N.D. Cal. 1992); *Apple Computer, Inc. v Microsoft Corp.* 35 F.3d 1435 (9th Cir. 1994).

[41] *Computer Associates International Inc v Altai, Inc.* 982 F.2d 693 (2d Cir. 1992); *Apple Computer, Inc. v Microsoft Corp.* 35 F.3d 1435 (9th Cir. 1994).

[42] *Computer Associates International Inc v Altai, Inc.* 982 F.2d 693 (2d Cir. 1992), 710-1.

Court authorities. These included a finding denying copyright for *de minimis* taking as being contrary to 'constitutional policies underlying the Copyright Act'.[43] Finally, the 9th Circuit Court of Appeals held that Apple and Microsoft's desktop interfaces were not identical and therefore copyright infringement could not be established in the circumstances.[44]

2.3.2 *Scope of Copyright Protection for Software in the European Union*

As discussed earlier in Section 2.2.1 and 2.2.2, the adoption of software copyright in the US prompted the EU to mandate that member states implement the Software Directive into national law. Along with the Information Society Directive (otherwise known as the 'Copyright Directive'), this Directive required EU member states to provide uniform copyright protection for software.[45] Specifically, the Software Directive requires protection of computer programs as 'literary works' under copyright law provided that those programs are 'original'.[46] However, the extent of originality and scope of copyright across the individual EU member jurisdictions is somewhat divided. This division emerges between the higher standard of originality for continental European jurisdictions, and a lower standard of originality found in Anglo-Saxon jurisdictions. For example, section 69a(2) of the German *Copyright Act 1965* (*Urheberrechtgesetz* or UrhG) states that the ideas underlying any element of a computer program are not entitled to protection under copyright law.[47] By contrast, in the United Kingdom, the *Copyright, Designs and Patents Act* was amended in 1997 to include software under the ambit of copyright protection.[48]

Nevertheless, as in the US, the question of what is copyrighted subject matter under the EU Software Directive is very much interpreted on a case by case basis.[49] Ostensibly, functionality is implicitly denied copyright protection under Article 1(2) of the Software

---

[43] *Computer Associates International Inc v Altai, Inc.* 982 F.2d 693 (2d Cir. 1992), 711; *Feist Publications, Inc. v Rural Telephone Service Co.* 499 U.S 340 (1991).

[44] *Apple Computer, Inc. v Microsoft Corp.* 35 F.3d 1435 (9th Cir. 1994).

[45] *Directive 91/250/EEC of 14 May 1991 on the Legal Protection of Computer Programs* [1991] *OJ L* 122/42; *Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights of the Information Society* [2001] *OJ L* 167/10; *Council Directive 2009/24/EC of 23 April 2009 on the Legal Protection of Computer Programs* [2009] *OJ L* 111/16.

[46] Articles 1(2) and 1(3) of the European Software Directive

[47] Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtgesetz) of 9.9.1965

[48] *Copyright, Designs and Patents Act 1988* (UK) section 3(2), amended by the *Copyright (Computer Programs) Regulations 1992 (UK) Copyright and Rights in Databases Regulations 1997* (UK)

[49] It should be noted that EU Directives do not create binding law by themselves. Instead, EU Directives provide specific minimum standards that member states must implement into national law. By contrast, EU Regulations are directly enforceable, even without implementation into national law. Nevertheless, EU case law establishes that where Union and member state law is consistent, Union law prevails. Accordingly, this Chapter will focus primarily on Union law.

Directive. However, Recital 11 of the Preamble notes that, to the extent to which 'programming logic, algorithms and languages' are composed of ideas and principles, copyright does not vest in the functional components of software. This definition suggests that outside programming logic, algorithms, and languages, copyright may vest in the functional aspects of software. This issue was raised in the *Bezpečnostní softwarová asociace - Svaz softwarové ochrany (BSA) v Ministry of Culture of the Czech Republic (BSA)* before the EU Court of Justice (CJEU). Specifically, the *BSA* case concerned the copyright eligibility of a graphical user interface. The CJEU held that copyright would vest in a work that was an author's 'own intellectual creation' (adopting the higher continental standard of copyright).[50] In part, the higher standard set by the CJEU may be attributed to the existence of an EU *sui generis* database right (which will be discussed in further detail below). However, the CJEU held that where 'different methods of implementing an idea are so limited that the idea and the expression becomes indissociable', the original authorship requirement shall not be met.[51] Although this doctrine differs from the US doctrine of attempting to discern the expression-functionality divide as opposed to the level of creativity, it nevertheless generates a similar result in permitting at least some copyright protection for functionality.[52]

The scope of EU copyright protection was once again considered in *SAS Institute v World Programming*,[53] which was heard before the English High Court and later the EU Court of Justice. SAS produced a programming language called Base SAS which allows users to run and develop their own statistical operations. World Programming then developed their own language which was backwards compatible with the Base SAS language. Crucially, unlike Google in *Oracle v Google*, World Programming did not directly copy the source code for Base SAS, but instead used a retail copy of SAS along with SAS user manuals to reverse engineer the functionality of Base SAS. However, both the High Court and the Court of Justice held that because World Programming had copied the functionality of Base SAS from manuals as well as other statistics textbooks, World Programming had not infringed the copyright of the protected aspects of Base SAS.[54] Whilst both courts denied SAS's particular cause of action, Shemtov nevertheless notes that each case left open the possibility that copyright could validly vest in the description of statistical functions.[55] This distinction would ultimately suggest that

---

[50] *Case C-393/09 Bezpečnostnı softwarová asociace - Svaz softwarové ochrany (BSA) v Ministry of Culture of the Czech Republic* [2010] ECR I–13971 para 42.

[51] *Case C-393/09 Bezpečnostnı softwarová asociace - Svaz softwarové ochrany (BSA) v Ministry of Culture of the Czech Republic* [2010] ECR I–13971 para 49.

[52] Noam Shemtov, *Beyond the Code: Protection of Non-Textual Features of Software* (Oxford University Press, 2017) 121-2.

[53] *SAS Institute Inc v World Programming Limited* [2011] RPC 1; *SAS Institute Inc v World Programming Ltd* [2013] EUECJ C–406/10.

[54] *SAS Institute Inc v World Programming Ltd* [2013] EUECJ C–406/10 para 46.

[55] Noam Shemtov, above n52, 123.

it may still be possible for copyright to vest in interfaces and other functional elements.

2.3.3 *The EU Sui Generis Database Right and Future EU Copyright Reform*

Relevant to the shift towards a higher standard of originality for copyright protection is the existence of an EU Database Directive, which creates a non-copyright, *sui generis* database right that vests in a compilation of facts.[56] EU member state courts have generally ruled against aggressive copyright protection for software (as discussed in the previous section). Accordingly, this *sui generis* right extends the scope of traditional copyright protection beyond derivative works to facts and data.[57] Rather than protection vesting in the actual facts themselves, the database right vests where there has been sufficient qualitative and quantitative effort in the selection or arrangement of the database.[58] The database directive will be infringed where a third party substantially reproduces parts of the database (in either a qualitative or quantitative fashion).[59] The database right does not necessarily impact on the question of open source licensing enforcement for software.[60] However, there remain unanswered questions as to the relationship between lawful use of a database and the use of text mining software to collect large amounts of data from databases.[61] In addition, any scientific or research reuse of a database must be strictly non-commercial and not involve the taking of substantial parts of the database.[62] In particular, Jerome Reichman, Tom Dedeurwaerdere and Paul Ulhir argue that existing database rights in scientific data could be easily infringed through the use of data mining software, which in turn would significantly stymie the use of such software.[63] This effect will be particularly pronounced as data science and big data analysis become increasingly intwined with scientific software development.[64] However, the next section will turn to address copyright protection for software in Australia and New Zealand.

---

[56] *Directive 96/9/EC on the Legal Protection of Databases* [1996] *OJ L* 77/20 No (028/11/EC).

[57] See Article 7 of the Database Directive.

[58] *Football Dataco Ltd and Others v Yahoo! UK Limited and Others* [2010] EUECJ C–604/10 paragraphs 47-52.

[59] *Forensic Telecommunications Services Ltd v West Yorkshire Police Anor* [2011] EWHC 2892 (Ch), paragraph 124.

[60] Simone Aliprandi, 'Open licensing and databases' (2012) 4(1) *International Free and Open Source Software Law Review* 5–18 10.

[61] Maarten Truyens and Patrick Van Eecke, 'Legal Aspects of Text Mining' (2014) 30(2) *Computer Law & Security Review* 153–170 164.

[62] *Forensic Telecommunications Services Ltd v West Yorkshire Police Anor* [2011] EWHC 2892 (Ch), paragraph 109-112.

[63] Jerome H. Reichman, Tom Dedeurwaerdere and Paul F. Uhlir, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge University Press, 2016) 339.

[64] Gregory Goth, 'Bringing Big Data to the Big Tent' (2015) 58(7) *Communications of the ACM* 17–19 19.

2.3.4 *Copyright Protection for Software in Australia and New Zealand*

In Australia the *Copyright Amendment Act 1984* (Cth) (the Amendment Act) amended section 10(1) of the *Copyright Act 1968* (Cth). This amendment confirmed that copyright protection is available for both the human readable aspects of computer programs as well as the object code generated after compilation.[65] This amendment was not only prompted by changes in the US, but also the case of *Apple v Computer Edge*,[66]. In this case, Justice Beaumont of the Federal Court interpreted the existing *Copyright Act 1968* (Cth) to suggest that copyright did not extend to object code (in the form of microcode). The Full Federal Court overturned Justice Beaumont's interpretation and extended copyright to object code[67]. However, the Full Federal Court's this interpretation was in turn overturned by the High Court which restored Justice Beaumont's original interpretation of the statute.[68] Nevertheless, the Australian government swiftly amended the *Copyright Act* to include both object and source code under the scope of copyright protection.[69]

Subsequent interpretations of the amended section 10(1) by the High Court in the *Autodesk v Dyason*[70] series of cases. These cases concerned whether the definition of computer program to include reserved words in a programming language as a form of 'set of instructions' that was eligible for copyright protection. The High Court further extended the scope of this definition in *Powerflex Services Pty Ltd v Data Access Corporation*,[71]. A majority held that a compression table generated using a Huffman compression algorithm[72] amounted to copyrighted subject matter according to a strict interpretation of the legislation. Anne Fitzgerald and Christina Cifuentes note that this broad interpretation of 'reserved words' would have the effect of rendering all copying of intermediate object code copyright infringement. Further, legitimate decompilation and reverse engineering would also amount to copyright infringement.[73] The *Copyright Act* was again amended in 2000 via the *Copyright Amendment (Digital Agenda) Act 2000* (Cth) (the '*Digital Agenda Act*'). This amendment was introduced in response to the passage of the WIPO Copyright Treaty, as well as the High

---

[65] *Copyright Act 1968* (Cth) section 10(1).

[66] *Apple Computer Inc v Computer Edge Pty Ltd* (1983) AIPC ¶90–121, 38,747-8.

[67] *Apple Computer Inc v Computer Edge Pty Ltd* (1984) AIPC ¶90–132, 38,828-9.

[68] *Computer Edge Pty Ltd v Apple Computer Inc* (1986) 161 CLR 171, 209.

[69] *Copyright Amendment (Computer Programs) Act 1984* (Cth).

[70] *Autodesk Inc v Dyason (No 1)* (1992) 173 CLR 330; *Autodesk Inc v Dyason (No 2)* (1993) 176 CLR 300.

[71] *Data Access Corp v Powerflex Services Pty Ltd* (1999) 202 CLR 1, para 125.

[72] The Huffman compression algorithm is an algorithm for the lossless encoding of data. See D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes" (1952) 40(9) *Proceedings of the IRE 1098–1101*

[73] Anne Fitzgerald and Christina Cifuentes, 'Accommodating Computer Software to Copyright Doctrine: Defining the Scope of Copyright Protection for Software' (2000) 11(2) *Journal of Law and Information Science* 224–253 245.

Court's suggestions regarding the need for statutory amendments to accommodate for software copyright.[74]

More recently in Australian case law concerning broader copyright issues, there has been a trend towards raising the threshold for originality and denying copyright infringement actions for *de minimis* taking. In particular, the decisions of the High Court in *Ice TV v Nine Network Australia*[75] and the Federal Court in *Telstra Corporation v Phone Directories Companies*[76] are relevant cases. Both cases reversed existing authority which extended copyright to compilations of material (such as television programming guides and phonebooks), thereby raising the requirements for copyright protection. The consequences of this shift for software copyright were demonstrated in the decision of the Full Federal Court in *Acohs Pty Ltd v Ucorp Pty Ltd*,[77]. In this case, hypertext output automatically generated from a database upon request was held not to constitute copyrightable subject matter. Andres Guadamuz argues that this decision significantly restricted the boundaries of copyrightable subject matter for software, particularly for automatically generated code.[78] Crucially, this standard of copyright eligibility creates a higher threshold for creativity, which may set Australia apart from other jurisdictions such as the US and the EU.

In New Zealand, the equivalent *Copyright Act 1994* (NZ) provides copyright protection for 'any form of notation or code, whether by hand or otherwise and regardless of the method' by which it is recorded. In New Zealand, the only decision on the scope of copyright protection for software is *International Business Machines Corp v Computer Imports Ltd*[79]. In this case, the New Zealand High Court followed the reasoning of the Australian High Court in *Apple v Computer Edge* in granting copyright to object code on the grounds that the object code was a reproduction of the original source code. However, neither Australian or New Zealand law has undergone significant evolution with respect to the scope of copyright protection for other more contentious pieces of software innovation, such as interfaces.[80].

---

[74] *Copyright Amendment (Digital Agenda) Act 2000* (Cth).

[75] *IceTV Pty Ltd v Nine Network Australia Pty Ltd* (2009) 239 CLR 459.

[76] *Telstra Corporation Ltd v Phone Directories Company Pty Ltd* [2010] FCA 44.

[77] *Acohs Pty Ltd v Ucorpo Pty Ltd* [2012] FCAFC 16.

[78] Andres Guadamuz, 'Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works' (2017) 2017(2) *Intellectual Property Quarterly* 169–186 184.

[79] *International Business Machines Corp v Computer Imports Ltd* [1989] 2 NZLR 395, 417.

[80] Susan Corbett, 'What if Object Code Had Been Excluded from Protection as a Literary Work in Copyright Law - A New Zealand Perspective' (2008) *Michigan State Law Review* 173–198 175.

2.4 OPEN SOURCE LICENSING AND COPYRIGHT PROTECTION

2.4.1 *The Birth and Enforcement of Open Source Licensing in the US*

Contemporaneously with the struggles on the scope of copyright for software, nascent free and open source software groups such as the League of Software Freedom (later the Free Software Foundation (FSF)) emerged. These groups were established in response to overly broad copyright claims on appearance and functionality. The FSF and others focused specifically on Apple's claims against Microsoft as examples of proprietary software developers attempting to enclose what were otherwise common place programming techniques.[81] Eventually, the FSF concluded that the only way to guarantee some protection for these techniques was to develop a copyright licensing system.[82] Richard Stallman, along with Eben Moglen, a lawyer affiliated with the FSF, drafted the first General Public Licence (GPL). In particular, the GPL prevents users of a GPL licensed package from redistributing that software or producing a derivative work of that software under the terms of another licence.[83]

Because the GPL places restrictions on the subsequent relicensing of licensed source code (otherwise known as *copyleft* restrictions), it and other software licences with copyleft terms have become known as *restrictive* open source licences, which must comply with the 'four essential freedoms'.[84] These restrictive licences can be contrasted with *permissive* open source licences. Permissive licences have generally emerged from either a particular research institute or laboratory which did not agree with the objectives of the GPL (such as the Berkeley or MIT licences) or from the development of a particular software package (such as the Apache licences).[85]

In contrast to copyleft licences, these licences do not mandate the relicensing of derivative works under the same licence.[86] The distinction between permissive and restrictive licences was formalised with the drafting of the Open Source Definition, which establishes ten criteria

---

[81] Richard M. Stallman, *Special Report: Apple's New Look and Feel* () GNU's Bulletin, vol. 1 no. 5 <https://www.gnu.org/bulletins/bull5.html#SEC9>.

[82] Christopher M. Kelty, *Two Bits: The Cultural Significance of Free Software* (Duke University Press, 2008) 179-208.

[83] Christopher M. Kelty, above n82, 208; Robert W. Gomulkiewicz, 'Conditions and Covenants in License Contracts: Tales from a Test of the Artistic License' (2008) 17(3) *Texas Intellectual Property Law Journal* 335–362 337.

[84] The four essential freedoms include: (0) the freedom to run the software for any purpose; (1) the freedom to study how the software works and to adapt it; (2) the freedom to redistribute copies of the software; (3) the freedom to improve the software and distribute those improvements publicly. See (Richard M. Stallman, *What is Free Software?* (12th June 2018) <https://www.gnu.org/philosophy/free-sw.en.html>)

[85] Luke McDonagh, 'Copyright, Contract and FOSS' in Noam Shemtov and Ian Walden (eds.), *Free and Open Source Software: Policy, Law and Practice* (Oxford University Press 2013)71–108 76-83.

[86] Robert W. Gomulkiewicz, 'Enforcement of Open Source Software Licenses: The MDY Trio's Inconvenient Complications' (2011) 14(1) *Yale Journal of Law and Technology* 106 114.

for open source licences to comply with.[87] Further, because each of these licences set different standards as to what can and cannot be reused in a derivative work, three key issues with respect to open source licences are:

1. whether they are validly formed;
2. whether they are enforceable by the copyright owner; and,
3. to *what extent* they apply to software.[88]

The first issue has often arisen in the context of proprietary licensing contracts that arise automatically as a result of the recipient using the product, otherwise known as 'shrinkwrap' or 'clickwrap' contracts. US courts have held that shrinkwrap licences are validly formed,[89] provided that the requisite elements for a contract are present[90] and 'reasonably conspicuous notice of the contract terms is brought to the user's attention' prior to use.[91] Shrinkwrap licensing is relevant to open source enforcement because many open source licences are drafted so that modification or distribution of the software amounts to acceptance of the licence.[92]

Copyleft licensing is designed to activate through use rather than formal offer and acceptance, with users being provided with formal notice before use. Accordingly, Jose Gonzalez de Alaiza Cardona notes that a user will not have accepted the terms if the open source licence is not explicitly brought to their attention.[93] To this end, the FSF (along with both commercial and non-commercial open source software distributors) publishes

---

[87] These ten criteria are (1) no restrictions on free distribution; (2) inclusion of source code with the program; (3) permission to develop modifications and derived works; (4) integrity of the author's source code; (5) no discrimination against persons or groups; (5) no discrimination against persons or groups; (6) no discrimination against fields of endeavour; (7) the rights under the licence must be capable of being executed without an additional licence; (8) the licence must not be specific to a product; (9) the licence must not restrict the use of other software; (10) the licence must be technology neutral. See (Open Source Initiative, *The Open Source Definition (Annotated) (Version 1.9)* (22nd March 2007) <https://opensource.org/osd>)

[88] Heather J. Meeker, 'Open Source and the Age of Enforcement' (2012) 4(2) *Hastings Science & Technology Law Journal* 267–290 273.

[89] *ProCD, Inc. v Zeidenberg* 86 F.3d 1447 (7th Cir. 1996).

[90] *Bragg v. Linden Research, Inc.* 487 F.Supp. 593 (E.D. Pa. 2007).

[91] *Specht v Netscape Communications Corp.* 306 F.3d 17 (2nd Cir. 2002).

[92] See GPL version 2.0 section 9 (Acceptance Not Required for Having Copies). 'You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work.'

[93] Dr. Jose J. Gonzalez de Alaiza Cardona, 'Open Source, Free Software, and Contractual Issues' (2007) 15(2) *Texas Intellectual Property Law Journal* 157 194-201.

recommendations on how developers should notify users of the existence of the GPL.[94] Nevertheless, outside the GPL, both Cardona and Robert Gomulkiewicz separately argue that an open source licensor may still rely on their rights available under copyright law. These rights will be available, even if an open source contract is not formed by reason of the licensee's use under copyright law.[95] Unfortunately for a prospective open source licensor, this argument is premised on open source licences being enforceable under copyright law. This argument largely depends on the structure of each licence. Further, the nature of an open source licensor's open non-exclusive rights raises the question of whether open source licence terms form part of an enforceable contract. This contract would therefore allow the licensor to control use of the licensed software and ensure that the source code remains openly available.[96]

Under US copyright law, if a licensor provides their work under a bare non-exclusive licence to multiple licensees with no conditions on use, each licensee has unfettered rights to use the software.[97] For US courts this lack of restriction has raised a vexing problem. In other words, where a non-exclusive licence is granted, what conditions or terms can give rise to an action in copyright infringement in addition to a breach of contract action.[98] If the terms of an open source licence are enforceable under copyright law as conditions, then the licensor can seek an injunction. This injunction could be either used to prevent use of the infringing software or require compliance with the licence. In addition, a licensor could also seek monetary remedies such as damages or an account of profits. However, many open source licences require the release of software without licensing fees. Therefore, an injunction to ensure compliance may be significantly more attractive than monetary remedies such as an account of profits so as to enforce the purpose of the licence. However, if an open source licence is held only to be a contractual licence, the licensor can only seek damages (rather than an injunction to control the use of the copyrighted work). Effectively, a limitation to damages would deny the open source licensor the ability to control how their software is used.[99]

---

[94] Free Software Foundation, *How to use GNU licenses for your own software* (March 2018) <https://www.gnu.org/licenses/gpl-howto.en.html>; Imed Hammouda et al., 'Open Source Legality Patterns: Architectural Design Decisions Motivated by Legal Concerns' (Paper presented at *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments* , 2010) 208.

[95] Robert W. Gomulkiewicz, above n86, 115.

[96] Dr. Jose J. Gonzalez de Alaiza Cardona, above n93, 200-1.

[97] *17 U.S. Code § 106 - Exclusive rights in copyrighted works*.

[98] Victoria Nemiah, 'License and Registration, Please: Using Copyright Conditions to Protect Free/Open Source Software' (2013) 3(2) *New York University Journal of Intellectual Property and Entertainment Law* 358 361.

[99] Robert W. Gomulkiewicz, above n83, 338-9; Andrés Guadamuz-González, 'The License/Contract Dichotomy in Open Licenses: a Comparative Analysis' (2009) 30(2) *University of La Verne Law Review* 101 104-5.

In *Graham v James*,[100] the Second Circuit of the US Court of Appeals dealt with a claim of software licence infringement. The case concerned Graham's failure to pay royalties and provide attribution to James, who had performed programming work for Graham pursuant to an implied oral contract. The Second Circuit held that a licence agreement had been formed in the circumstances and that there had been a failure to pay royalties and provide attribution. However, the Second Circuit also held that attribution and royalties were not conditions on use of the software under the contractual licence. In particular, the Second Circuit noted that James had given the program for use by Graham before any royalties were paid. Graham's failure to comply with each of the requirements described above did not amount to a breach of a condition on use under copyright. Therefore, James could not bring an action for breach of copyright against Graham. James was therefore left with contractual damages as a remedy as opposed to an injunction to prevent Graham's unfettered use of the software.[101] *Graham v James* is relevant to open source licensing because the requirement for attribution is a minimum requirement for all open source licences. However, *Graham v James* would suggest that failure to provide attribution is not a breach of a condition of use. Therefore, this failure to attribute does not give rise to an action in copyright infringement.[102]

The issue of which conditions, when violated, would support an action for copyright infringement was again considered by the US District Court of California in *Sun Microsystems v Microsoft*.[103] Sun and Microsoft had entered into a written Technology Distribution and Licence Agreement (TLDA) with respect to Sun's Java Virtual Machine (JVM) software. JVM allowed the same program written in the Java programming language to be run on multiple operating systems, rather than a different program being programmed separately for each operating system.[104] In other words, a JVM could run the same program on Microsoft Windows, Apple's Macintosh OS or any other operating system. The TLDA granted Microsoft non-exclusive distribution and development rights to modify JVM source code in exchange for Microsoft complying with Sun's interoperability requirements.

Sun alleged that Microsoft had breached this copyright licence by distributing software that would only work with Microsoft Windows and did not satisfy the compatibility requirements in the TLDA. Microsoft's objective in failing to comply with the TLDA was to 'kill cross platform Java compatibility', therefore protecting their monopoly in the operating system market.[105]

---

[100] *Graham v James* 144 F.3d 229 (2d Cir. 1998).

[101] *Graham v James* 144 F.3d 229 (2d Cir. 1998).

[102] Victoria Nemiah, above n98, 368.

[103] *Sun Microsystems Inc. v Microsoft Corporation* (ND. Cal., No No. C 97-20884RMWPVT, 2000 WL 33223397, May 2000).

[104] Tim Lindholm and Frank Yellin, *The Java Virtual Machine Specification* (Addison-Wesley Longman Publishing Co., Inc., 2nded, 1999).

[105] *Memorandum Of The United States In Support Of Motion For Preliminary Injunction?: U.S. V. Microsoft*

However, the District Court reversed its earlier decision,[106] and held, based on the construction of the TLDA, that the compatibility requirements were not mentioned in the terms actually granting Microsoft the distribution and development rights. Accordingly, the District Court characterised the compatibility requirements as contractual terms, the breach of which could not give rise to copyright infringement.[107] This result was problematic for Sun, as the contractual remedies that were available to them were limited to revocation of the TLDA (as opposed to damages or an injunction that would prevent Microsoft from developing a variant of Java).[108]

2.4.2 *Jacobsen v Katzer as a Landmark Case in Open Source Licence Enforcement*

These cases seemingly rejected the notion that failure to provide attribution amounted to a breach of condition. However, it was not until the *Jacobsen v Katzer* series of cases that these issues were considered for open source licences.[109] Jacobsen, the plaintiff, had released a software interface to control model trains under the Artistic Licence, a restrictive open source licence.[110] An employee of Katzer, the defendant, subsequently modified Jacobsen's software and sold it commercially under a proprietary licence. Katzer then obtained a utility patent for the method of interaction with a train chip underlying the software. In response, Jacobsen brought an action against Katzer for copyright infringement for breach of the Artistic Licence. Complicating matters further was the uncertain legal status of the Artistic Licence, which the FSF described as being 'too vague' to be a restrictive open source licence. At first instance, the US District Court for the Northern District of California held that the violation of the licence agreement amounted to a prima facie breach of contract. However, the District Court denied the claim for copyright infringement on the grounds the Artistic Licence was 'intentionally broad'.[111] Specifically, the Court held the Artistic Licence allowed any member of the public 'the right to use and distribute the [material] in a more-or-less customary fashion, plus the right to make reasonable accommodations', which in turn precluded the licensor's right to sue for copyright infringement.[112]

Corporation | ATR | Department of Justice <http://www.justice.gov/atr/memorandum-united-states-support-motion-preliminary-injunction-us-v-microsoft-corporation>.

[106] *Sun Microsystems, Inc. v Microsoft Corp.* 21 F.Supp. 1109 (1998).

[107] *Sun Microsystems Inc. v Microsoft Corporation* (ND. Cal., No No. C 97-20884RMWPVT, 2000 WL 33223397, May 2000).

[108] Victoria Nemiah, above n98, 373-6.

[109] *Jacobsen v Katzer* (N.D. Cal., No No. 06-CV-01905 JSW, 2007 WL 2358628, December 2007); *Jacobsen v Katzer* 535 F.3d 1373 (Fed. Cir. 2008).

[110] Various Licences and Comments About Them <https://www.gnu.org/licenses/license-list.htmlNonFreeSoftwareLicense>

[111] *Jacobsen v Katzer* (N.D. Cal., No No. 06-CV-01905 JSW, 2007 WL 2358628, December 2007).

[112] *Jacobsen v Katzer* (N.D. Cal., No No. 06-CV-01905 JSW, 2007 WL 2358628, December 2007).

As Heather Meeker notes, the District Court's rejection of Jacobsen's copyright claim would preclude any copyright action for breach of an open source licence term. The lack of a copyright claim would flow from the fact that open source terms were not conditions on use.[113] Further, as Bryan Carson notes, under Californian law the breach of contract action that remained available to Jacobsen did not give rise to a presumption of an injunction. The lack of an injunction would leave permissive open source licensors without a mechanism to enforce attribution. Likewise, the lack of an injunction for breach of a restrictive open source licence would deny licensors the ability to ensure that the 'copyleft' terms of the licence are complied with. The absence of an injunction would therefore prevent derivatives from being licensed under the same open source licence.[114] Consequently, open source advocates were relieved when the District Court's decision was overturned and reversed by the Court of Appeals for the Federal Circuit.[115]

The analysis of the Federal Circuit's decision can be split between the consideration of the contractual and copyright aspects of the Artistic Licence. With respect to the contractual elements, the Federal Circuit provided considerable theoretical support for the non-monetary economic benefits of open source. These non-monetary benefits included attribution rights flowing to Jacobsen from users utilising his software and therefore seeking his services for other software projects.[116] Accordingly, although the District Court had already held the Artistic Licence was a valid contract (without providing an explicit justification), the Federal Circuit expressly recognised that the benefits from using open source licensed software would be satisfactory consideration to support a contract despite being non-monetary in nature:[117]

> Traditionally, copyright owners sold their copyrighted material in exchange for money. The lack of money changing hands in open source licensing should not be presumed to mean that there is no economic consideration, however. There are substantial benefits, including economic benefits, to the creation and distribution of copyrighted works under public licences that range far beyond traditional licence royalties. For example, program creators may generate market share for their programs by providing certain components free of charge. Similarly, a programmer or company may increase its national or international reputation by incubating open source projects. Improvement to a product can come rapidly and free of charge from an expert not even known to the copyright holder.

---

[113] Heather J. Meeker, above n88, 275.

[114] Bryan Carson, 'Legally Speaking – The Legality of Open Source Part II: Jacobsen v. Katzer' (2013) 21(1) *Against the Grain* 59.

[115] *Jacobsen v Katzer* 535 F.3d 1373 (Fed. Cir. 2008).

[116] *Jacobsen v Katzer* 535 F.3d 1373 (Fed. Cir. 2008), 1379.

[117] David Ferrance, 'Economic Interests and Jacobsen v. Katzer: Why Open Source Software Deserves Protection under Copyright Law' (2010) 58(4) *Journal of the Copyright Society of the USA* 819 566.

The court continued:

> It is logical that as the Software improved, more end-users used [Jacobsen's] Software, thereby increasing [sic] recognition in his profession and the likelihood that the software would be improved even further.[118]

The District Court had already held the Artistic Licence was a valid contract. However, the Federal Circuit expressly recognised the benefits from using open source licensed software, including the recognition to the original developer and the potential from improvement through use. These benefits would be satisfactory consideration to support a contract despite being non-monetary in nature.[119] On the issue of copyright infringement, the Federal Circuit framed the question as whether Katzer's actions (as a licensee) fell outside the scope of conditions of use provided by the Artistic Licence. This issue in turn raised the question of whether Jacobsen (as a licensor) was entitled to sue for copyright infringement for Katzer's breaches of the Artistic Licence.[120] The Federal Circuit then examined the terms of the Artistic Licence. The Federal Circuit noted that the requirement to include a notice on what files were changed was explicitly stated to be a condition for use of the software. In particular, the Federal Circuit held these conditions were 'clear and necessary to accomplish the objectives of open source software development'.[121] The fact there were no explicit economic benefits flowing from the software's distribution to Jacobsen was not enough to preclude Jacobsen's right to copyright protection under the Artistic Licence. The Federal Circuit therefore held that Katzer's breach of the Artistic Licence entitled Jacobsen to injunctive relief, noting that without injunctive relief the licence would be rendered useless.[122] Despite the fact that the Artistic Licence lacked the explicit relicensing requirements, the attribution requirement was still found to be enforceable.[123]

2.4.3 *MDY Industries v Blizzard and the Contraction of Open Source Licence Enforcement*

Unfortunately, the decision of *MDY Industries v Blizzard*[124] has cast doubt on the enforceability of open source licences. In this case, Blizzard required users to accept a proprietary EULA before they could play Blizzard's multiplayer online computer game *World*

---

[118] *Jacobsen v Katzer* 535 F.3d 1373 (Fed. Cir. 2008), 1379.

[119] David Ferrance, above n117 566.

[120] *Jacobsen v Katzer* 535 F.3d 1373 (Fed. Cir. 2008), 1376.

[121] *Jacobsen v Katzer* 535 F.3d 1373 (Fed. Cir. 2008), 1376, 1381.

[122] *Jacobsen v Katzer* 535 F.3d 1373 (Fed. Cir. 2008), 1382.

[123] Victoria Nemiah, above n98, 381.

[124] *MDY Industries, LLC v Blizzard Entertainment* 629 F.3d 928 (9th. Cir. 2010).

of Warcraft. MDY Industries developed software that would allow players to automate earlier levels of the game. Blizzard responded by redesigning World of Warcraft to detect third party software being used by the game client. When MDY attempted to avoid detection, Blizzard sued for copyright infringement on the grounds that MDY's software amounted to a breach of the EULA. In response, MDY argued there was only a breach of contract, as the terms preventing automated software were not conditions controlling use under the copyright licence.[125] The Ninth Circuit held that a breach of licence would constitute copyright infringement if two conditions were satisfied. First, the infringement must have exceeded the licence's scope. Secondly, the infringement must have implicated one of the licensor's exclusive statutory rights.[126] With respect to the second criterion, the Ninth Circuit defined an exclusive right to be narrowly limited to unlawful reproduction or distribution, noting that contractual limitations could be much broader. On the facts, the Ninth Circuit concluded that the term of the EULA prohibiting automated software did not have sufficient proximity to Blizzard's exclusive statutory rights to give rise to copyright infringement.[127]

At present, academic opinion is mixed as to the effect of *MDY v Blizzard*. Significant debate focuses on whether open source licence requirements of distribution of source code and relicensing of derivative works would have a sufficient nexus to the licensor's exclusive rights at copyright law. On the one hand, as Victoria Nemiah notes, *MDY v Blizzard* places onus on the open source licence drafters to ensure that terms are explicitly characterised as conditions on use.[128] In addition, Curt Blake and Joseph Probst suggests that attention to detail with licence drafting is particularly important for restrictive licences. These restrictive licences are more likely to give rise to actions for copyright infringement or breach of contract because of the contentious nature of the copyleft relicensing requirements.[129] In contrast, Peter Maggs notes that the divergent results in *Jacobsen v Katzer* and *MDY Industries v Blizzard* could be attributable to different interpretations of state contract law. Specifically, *Jacobsen v Katzer* related to Californian state contract law, whereas *MDY Industries v Blizzard* concerned Arizonan state contract law.[130] Therefore, the question of open source licence enforcement in the US and the distinction between these two decisions will not be resolved by an overarching Supreme Court decision.

---

[125] *MDY Industries, LLC v Blizzard Entertainment* 629 F.3d 928 (9th. Cir. 2010).

[126] *MDY Industries, LLC v Blizzard Entertainment* 629 F.3d 928 (9th. Cir. 2010), 940.

[127] *MDY Industries, LLC v Blizzard Entertainment* 629 F.3d 928 (9th. Cir. 2010), 942.

[128] Victoria Nemiah, above n98, 384.

[129] Curt Blake and Joseph Probst, 'Loaded Question: Examining Loadable Kernel Modules under the General Public License v2' (2011) 7(3) *Washington Journal of Law, Technology & Arts* 265–294 287.

[130] Peter B. Maggs, 'License Contracts, Free Software and Creative Commons in the United States' (2014) 62(suppl_1) *The American Journal of Comparative Law* 407–423 411.

2.4.4 *Open Source Licensing in the EU*

There has been support for the use of open source licences and open data strategies by the EU at a supranational level. However, despite emphasising the importance of interoperability, neither the Software Directive or the Information Society Directive contain an explicit definition of derivative works. The absence of a specific definition for derivative works undermines the potential for open source licences, which only extend as far as derivative works.[131] Whilst explicitly recognising the enforceability of open source licences is one solution, at the national level only a minority of European jurisdictions support alternative licences under copyright law. These include the Czech Republic, France, Germany, Italy and Portugal. Case law on the enforcement of open source licences is limited to Belgium, France, Germany, the Netherlands and Spain.[132] This section considers the two jurisdictions, France and Germany, with the most substantive case law. In France, the *Code de la propriété intellectuelle* was amended in 2006 to introduce an article explicitly supporting licensing 'free of consideration'.[133] Although this article has no enforceability per se, the Paris Court of Appeals case of *Educaffix v CNRS* established that software developers should be aware of the existence of open source licences.[134] Further, the subsequent Paris Court of Appeals case of *EDU 4*[135] established that the GPL was enforceable when incorporated into a contract for the sale of software.

In addition, German precedent consistently and directly supports the enforceability of open source licences as contractual copyright licences. The first German open source licence case concerned Harald Welte, a German computer programmer who brought a series of actions regarding *iptables*. Welte had helped developed *iptables*, which was licensed under the GPL. Welte alleged that source code from his software had been incorporated into proprietary embedded software that ran on wireless routers.[136] Welte's actions reflect concerns in the open source community about the use of hardware restrictions to avoid compliance with open source licence conditions, otherwise known as 'tivoisation'.[137] The neologism 'tivosation'

---

[131] Lothar Determann, 'Dangerous Liaisons - Software Combinations as Derivative Works - Distribution, Installation, and Execution of Linked Programs under Copyright Law, Commercial Licenses, and the GPL' (2006) 21(4) *Berkeley Technology Law Journal* 1421 1469.

[132] Axel Metzger and Stefan Hennigs, 'General Report' in Axel Metzger (ed.), *Free and Open Source Software (FOSS) and other Alternative License Models: A Comparative Analysis* (Springer 2015) 6-11.

[133] *Code de la propriété intellectuelle 1992* (French Intellectual Property Code) Article L. 122-7-1.

[134] Tribunal de grande instance de Paris (CA Paris),22nd January 2008 Gaz. Pal. no 22.

[135] Cour d'appel de Paris (CA Paris),16th September 2009 RG no 01/24298.

[136] *Welte v Sitecom (Open Source - effectiveness of GPL)*, Landgericht (District Court München I) 19th May 2004 reported in 2004 MMR 693; *Welte v D-Link*, Landgericht Frankfurt (District Court of Frankfurt) 22nd September 2006 reported in 2006 CR 729; *Welte v Skype*, Landgericht (District Court München I) 12th July 2007 reported in 2007 CR 57.

[137] James E. Bottomley, *Kernel developers' position on GPLv3 [LWN.net]* (22nd September 2006) <http://lwn.net/

refers to TiVo, a digital video recorder that uses GPL version 2 licensed software but places hardware restrictions on the device in order to prevent compliance with GPL version 2. The District Court of Munich considered whether the terms requiring relicensing under the GPL and attribution for the original author contradicted either the German Copyright Act or the German Civil Code (Bürgerliches Gesetzbuch or BGB). The court first turned to the issue of contractual validity, holding the terms of the GPL were contractual terms. The court then addressed the specific question of whether the GPL unfairly disadvantaged the licensee or illegitimately restricted use of content. The court held that the automatic termination could be characterised as a 'condition subsequent' that did not unfairly restrict the distribution of copyrighted content. Accordingly, the court held that the GPL was a valid copyright licence and therefore that Welte did not waive his rights to sue through the distribution of the software under the licence. Therefore, Welte could seek an injunction to prevent the distribution of his software outside the terms of the GPL.[138] The court gave particular emphasis to the negative effects on open source development should the GPL be held to be invalid. This finding mirrored the Federal Circuit's discussion in *Jacobsen* regarding the importance of supporting open source development.[139]

There have been subsequent attempts at enforcing the GPL in Germany. In the *WISO Mein Büro*[140] case, the plaintiff successfully instituted an action for the breach of the Lesser GPL licence. Less successfully, Patrick McHardy, a former member of the *iptables* development team, attempted to bring a number of actions against potential GPL infringers, but sought damages rather than licence compliance. This action was considered contrary to the principles of GPL enforcement established by the Software Freedom Law Conservancy, leading to McHardy's suspension from the *iptables* team.[141] Further, an action brought by Christoph Hellwig against VMWare for GPL licence infringement was also rejected in 2017. Hellwig's case failed on the grounds that Hellwig did not identify the parts of GPL licensed code that VMWare had allegedly infringed.[142] Both of these cases demonstrate that even where there is precedent supporting the enforcement of open source licences, enforcement action must be

---

Articles/200422/>.

[138] *Welte v Sitecom (Open Source - effectiveness of GPL)*, Landgericht (District Court München I) 19th May 2004 reported in 2004 MMR 693.

[139] Julian P Hoppner, 'The GPL Prevails: An Analysis of the First-Ever Court Decision on the Validity and Effectivity of the GPL' (2004) *SCRIPT-ed* 628–635.

[140] *adhoc data service GmbH v Buhl Data Service GmbH (WISO Mein Büro)*, Landgericht (District Court of Bochum) 20th January 2011 reported in 2011 CR 29.

[141] Richard Fontana, 'Seven Notable Legal Developments In Open Source In 2016' (2017) 8(1) *International Free and Open Source Software Law Review* 59–65 60-1.

[142] Scott K. Peterson, *GPL Enforcement Action in Hellwig v. VMware Dismissed, With an Appeal Expected* (11th August 2016) opensource.com <https://opensource.com/law/16/8/gpl-enforcement-action-hellwig-v-vmware>.

targeted correctly to prevent wasting resources with frivolous actions.[143] Subsequent chapters of this thesis will explore not only the different licence enforcement strategies that bioinformatics projects rely on, but the rules that govern how these licences are enforced.

2.4.5 *Open Source Licensing in Australia and New Zealand*

In contrast to the US, there has been no case law on the enforcement of open source licences in Australia and New Zealand. This section considers the limited case law on the enforcement of shareware and other non-monetary software licences. Whilst these licences do not have equivalent open source licensing provisions, they are 'free' and still operate without monetary consideration. In *Trumpet Software v OzEmail*,[144] the plaintiff (Trumpet Software) brought a copyright infringement action against the defendant (OzEmail) for distributing and modifying Trumpet's shareware software without permission. Justice Heerey held that the distribution and modification of a substantial amount of Trumpet's software amounted to a breach of the shareware licence. Justice Heerey further held the shareware licence was a valid copyright licence that permitted OzEmail to distribute Trumpet's software but which Trumpet could revoke at any time. However, Justice Heerey also held that there was no consideration moving from the licensee (OzEmail) to the licensor (Trumpet). Accordingly, Trumpet had no other remedies to rely upon aside from revocation within a reasonable period of time.

Justice Heerey's decision in *Trumpet* has both positive and negative implications for open source software licence enforcement in Australia. On one hand, His Honour recognised the distribution of software without charge does not mean that the licensor waives their copyright over software.[145] However, His Honour refused to recognise that the non-monetary benefits associated with open source licences could constitute sufficient consideration to give rise to a contract. As open source licences do not require express monetary consideration, this norm would deny open source licensors contractual remedies resulting from the licensee acting outside the scope of the licence. In obiter (as Trumpet had already revoked the licence) Justice Heerey was only prepared to imply that the shareware licence was a bare licence that gave a right to revocation. This line of reasoning followed the earlier case of *Computermate v Ozi-Soft*[146] where the Federal Court held that a bare licence could exist without a contract. Further, even if that licence were gratuitous and non-exclusive in nature, it could still act as a

---

[143] Siobhán O'Mahony, 'Guarding the Commons: How Community Managed Software Projects Protect their Work' (2003) 32(7) *Research Policy* 1179–1198 1183.

[144] *Trumpet Software Pty Ltd v OzEmail* (1996) 34 IPR 481.

[145] Cristina Cifuentes and Anne Fitzgerald, 'Copyright in Shareware Software Distributed on the Internet—the Trumpet Winsock Case' (Paper presented at *Proceedings of the 19th International Conference on Software Engineering* , 1997).

[146] *Computermate Products (Aust) Pty Ltd v Ozi-Soft Pty Ltd* (1988) 12 IPR 487.

defence against copyright infringement. As a result, the licensor in *Computermate* could only rely on reasonable revocation when faced with a breach of the licence. *Computermate* must be considered in context, as it concerned an importation licence for the distribution of software without restrictions on how the importer licensee could deal with the software. As open source licences contain additional requirements for redistribution and attribution, they may be characterised as more than bare distribution licences.

Considering Justice Heerey's decision, Ben Giles argues that to avoid liability, open source developers should compromise and release software with a nominal licence fee to satisfy contractual requirements.[147] Nevertheless, despite Justice Heerey's characterisation of contract law there is UK and Australian precedent recognising non-monetary consideration to be sufficient.[148] In concert with the explicit recognition of non-monetary benefits from open source licences in *Jacobsen*, a future Australian court may overturn Justice Heerey's decision and decide that the benefits from open source licensing are sufficient to satisfy the requirements for consideration. This argument may be supported by the fact that governmental agencies in Australia are increasingly embracing open source style licensing for public sector data.[149] However, for the moment the enforceability of open source licences in Australia remains uncertain and will largely be dependent on how these licences are drafted.[150]

Another divergent area of copyright law that might affect the enforcement of open source licences is the difference between the US standard of 'derivative work'[151] and the Australian standard of 'adaptation',[152] which was inherited from UK copyright law.[153] As discussed in Section 2.4.1 above, Moglen and Stallman relied on the derivative work right to control how developers would subsequently license their own versions of GPL licensed software.[154]

---

[147] Ben Giles, 'Consideration and the Open Source Agreement' (2002) 49(September) *NSW Society for Computers and the Law*.

[148] *Chappell & Co Ltd v. Nestle Co Ltd* [1960] AC 87; *Coastle Recycled Cooking Oils Pty Ltd v Innovative Business Action and Strategies Pty Ltd* [2007] NSWSC 831; *The Movie Network Channels Pty Ltd v Optus Vision Pty Ltd* [2010] NSWCA 111.

[149] Judith Bannister, 'Open Government: From Crown Copyright to the Creative Commons and Culture Change Forum: Intellectual Property Law in Australia' (2011) 34(3) *University of New South Wales Law Journal* 1080–1103 1098-1103.

[150] James Scheibner and Dianne Nicol, 'Do Software Patents Inhibit Open Source Licensing in Australia?' (2015) 25(4) *Australian Intellectual Property Journal* 198.

[151] *17 U.S. Code § 101 - Definitions* 'A work consisting of editorial revisions, annotations, elaborations, or other modifications which, as a whole, represent an original work of authorship, is a "derivative work"'.

[152] *Copyright Act 1968* (Cth) 'section 10 - "adaptation" means in relation to literary work being a computer program - a version of the work (whether or not in the language, code or notation in which the work was originally expressed) not being a reproduction of the work'; *Copyright Act 1994* (NZ) section 10(1)(b).

[153] Patrick R. Goold, 'Why the U.K. Adaptation Right Is Superior to the U.S. Derivative Work Right' (2013) 92(4) *Nebraska Law Review* 843–896 846.

[154] Brian Fitzgerald and Nic Suzor, 'Legal Issues for the Use of Free and Open Source Software in Government'

Nevertheless, it is unclear whether the Australian adaptation right extends further to include the reuse of libraries from an open source project, with the Federal Court of Australia in *Coogi Australia Pty Ltd v Hysport International Pty Ltd*[155]. In this case, the Federal Court contrasting the compilation of source code (as an adaptation) with the reproduction of a functionally equivalent computer program (as not an adaptation).[156] Accordingly, this definition of 'adaptation' under Australian copyright law does not assist with the determination of how far the viral licensing requirements under the GPL extend.[157]

2.5 EXCEPTIONS TO COPYRIGHT INFRINGEMENT FOR SOFTWARE

2.5.1 *Exceptions to Copyright Infringement for Software in the US*

In *Sega Enterprises Ltd v Accolade Inc* (*Sega Enterprises Ltd*),[158] the 9th Circuit Court of Appeals considered the implications of attempting to reverse engineer the functionality of a computer program. The case concerned video game publisher Accolade's attempts to reverse engineer a Genesis video game console produced by Sega so that they could produce games for the console without having to pay exorbitant licence fees to Sega.[159] This reverse engineering process involved Accolade copying and decompiling the object code on a Sega console and then rewriting the source code for their own software so that the resultant code was compatible with a Sega console. Sega sued Accolade for copyright infringement, but Accolade counter argued that their decompilation was fair use. The 9th Circuit Court of Appeals confirmed that in this instance, although Accolade had breached copyright by decompiling and copying the object code, the replication of functionality amounted to fair use under the amended US copyright law.[160] In applying the four fair use factors described in Section 2.2.2, the 9th Circuit noted that Accolade's decompilation as an act of copying was limited to the functional aspects of Sega's software, and that their decompilation was sufficiently transformative (insofar that it was not envisaged by Sega as the original creator) to amount to fair use.[161] This application of the fair use defence can be contrasted with the contemporaneous case of *Atari Games Corp v Nintendo America Inc*.[162] In this case, Atari requested a copy of Nintendo's source code

---

(2005) 29(2) *Melbourne University Law Review* 412 416.

[155] *Coogi Australia Pty Ltd v Hysport International Pty Ltd* (1998) 41 IPR 593.

[156] *Coogi Australia Pty Ltd v Hysport International Pty Ltd* (1998) 41 IPR 593, paragraph 621-4.

[157] Mary Huang, 'The Problems of Openness – Effective Regulation of Open Source Software' (2012) 1(1) *International Journal of Technology Policy and Law* 48–68 55-6.

[158] *Sega Enterprises Ltd. v Accolade, Inc.* 977 F.2d 1510 (9th Cir. 1992).

[159] *Sega Enterprises Ltd. v Accolade, Inc.* 977 F.2d 1510 (9th Cir. 1992), 1514-5.

[160] *Sega Enterprises Ltd. v Accolade, Inc.* 977 F.2d 1510 (9th Cir. 1992), 1519-20.

[161] *Sega Enterprises Ltd. v Accolade, Inc.* 977 F.2d 1510 (9th Cir. 1992), 1526.

[162] *Atari Games Corp v Nintendo of America Inc* 975 F.2d 832 (Fed. Cir. 1992), 843.

from the US Copyright Office to prepare for an alleged copyright infringement action, but were in fact using this source code in order to assist with the decompilation of Nintendo's object code. Nevertheless, both *Sega Enterprises* and *Nintendo* exist as authority for the decompilation and reverse engineering of the functionality of copyrighted software being broadly permissible under fair use.[163]

Whilst this precedent established a broad scope for creating interoperable software, developers began to rely on both technological protection measures (TPMs) and contract law (such as End User Licence Agreements (EULAs)). The use of TPMS and EULAs was designed to prevent third parties from decompiling object code.[164] These mechanisms were also reinforced in the US by the passage of the *Digital Millennium Copyright Act* (the *DMCA*) in 1998. The DMCA was warranted due to the US's ascension to both the World Intellectual Property Organisation (WIPO) Copyright and Performances and Phonograms Treaties (the 'WIPO Copyright Treaty').[165] Amongst other things, the WIPO Copyright Treaty permitted rights holders to include TPMs along with their object code to prevent infringement except in narrow reverse engineering circumstances.[166] Finally, many US software developers began to pursue patent protection for software. The implications of this shift will be discussed in further detail in Chapter Three. Effectively, the combination of restrictive licensing and copyright protection through TPMs created extensive intermeshed protection between copyright and trade secrecy with respect to the underlying software source code.[167]

As mentioned previously, *MDY Industries v Blizzard* is the most recent adjudication in the US to explicitly consider the enforceability of software licence copyright conditions. However, the more recent dispute between Oracle and Google considered issues surrounding the enforcement of open source licences, albeit in the context of software interoperability.[168] This dispute concerned Google's use of Java application programming interfaces in their Android operating system. Some of these APIs were licensed under the GPL, but others were released under dedicated proprietary licences.[169] At the time, Sun Microsystems owned the

---

[163] Lothar Determann, above n131, 1446.

[164] Noam Shemtov, above n4, 104-7.

[165] *World Intellectual Property Organisation Copyright Treaty*, opened for signature 20th December 1996, 36 ILM 65 (entered into force 6th March 2002); *WIPO Performances and Phonograms Treaty*, opened for signature 20th December 1996, 36 ILM 76 (entered into force 20th May 2002).

[166] *Digital Millennium Copyright Act 17 USC § 1201(a) 1998*; Lothar Determann and David Nimmer, 'Software Copyright's Oracle from the Cloud' (2014) 30(1) *Berkeley Technology Law Journal* 161 204.

[167] Robert Tomkowicz, *Intellectual Property Overlaps: Theory, Strategies, and Solutions* (Routledge, 2013) 204-6.

[168] *Oracle America, Inc. v Google Inc.* 872 F.Supp.2d 974 (Dist. Court, ND California 2012), 70; *Oracle America, Inc. v Google Inc.* 750 F.3d 1339 (Fed. Cir. 2014); Pamela Samuelson, 'Evolving Conceptions of Copyright Subject Manner' (2016) 78(1) *University of Pittsburgh Law Review* 17–94.

[169] *Oracle America, Inc. v Google Inc.* 750 F.3d 1339 (Fed. Cir. 2014), 1350.

rights to these APIs. However, in contrast to Sun's dispute with Microsoft, where there was a licensing agreement, negotiations between the two parties for Google to include APIs in their operating system proved inconclusive.[170] Google then independently copied the structure, sequence and organisation (SSO) of Sun's APIs (including variable names and command operations). Whilst Sun was aware of Google's reimplementation, Sun was bought out by Oracle in 2010, who immediately sued Google for copyright infringement. Oracle conceded that Google had only copied the SSO for the APIs (unlike the interface designs contested in *Apple v Microsoft* and *Computer Associates v Altai*). However, Oracle argued the APIs in question required a substantial degree of creativity to develop.[171] Google denied that copyright vested in the structure, sequence and operation of Oracle's APIs in question. In the alternative, Google argued that even if copyright did vest in the APIs, their reimplementation was protected under the fair use exception.

At first instance, Judge Alsup of the District Court of California dismissed Oracle's copyright claim. Judge Alsup held that 'as long as the specific code is different', another developer could write software to carry out the same functions as the Java API.[172] However, on appeal, the Federal Circuit upheld Oracle's copyright claim, holding that Judge Alsup erred in finding that the functionality in the APIs was not protected by copyright.[173] Crucially, the Federal Circuit held that copyright would vest in the APIs even though they were functional works, or else no copyright could vest in any software. Instead, the Federal Circuit held that copyright could vest in a functional work, so long as there were multiple ways of expressing that work.[174] Although the Supreme Court refused leave to hear an appeal, it nevertheless remitted the decision to the District Court to determine whether Google's use of the APIs amounted to fair use. Judge Alsup again held that Google had only reimplemented the APIs that it needed for smartphone development.[175] Accordingly, Judge Alsup Google's reverse engineering was sufficiently transformative to amount to fair use.[176] Oracle again appealed to the Federal Circuit, which reversed Judge Alsup's decision on the grounds that Google's copying of Sun's APIs was not protected by fair use. Despite the fact that Google released their APIs without charge, the Federal Circuit reasoned that these APIs formed part of a broader Android software ecosystem. The Federal Circuit concluded that these APIs allowed

---

[170] Pamela Samuelson, 'Oracle V. Google: Are APIs Copyrightable?' (2012) 55(11) *Communications of the ACM* 25–27 25.

[171] *Oracle America, Inc. v Google Inc.* 872 F.Supp.2d 974 (Dist. Court, ND California 2012), 975-6; Pamela Samuelson, above n170, 26.

[172] *Oracle America, Inc. v Google Inc.* 872 F.Supp.2d 974 (Dist. Court, ND California 2012), 998.

[173] *Oracle America, Inc. v Google Inc.* 750 F.3d 1339 (Fed. Cir. 2014), 1359.

[174] *Oracle America, Inc. v Google Inc.* 750 F.3d 1339 (Fed. Cir. 2014), 1367.

[175] By contrast, Sun's APIs were originally developed for Java applications on desktop and laptop computers.

[176] *Oracle America, Inc. v Google Inc.* (Dist. Court. ND California, No No. C 10-03561 WHA, May 2016) 7-10.

Google to compete directly with Oracle in the market for mobile devices (despite the fact that Oracle did not release mobile devices at the time).[177]

In neither decision did the Federal Circuit touch on the obiter from Judge Alsup's judgement[178] which opined software functionality was better suited by patent protection.[179] In addition, the Supreme Court has declined to review the Federal Circuit's decision on copyright protection.[180] Nevertheless, Google will almost certainly appeal on the question of fair use. In light of the restricted scope of patent protection for software in the US which is discussed in Chapter Three, the consequences of the *Oracle v Google* decision for software copyright protection are somewhat uncertain. Richard Fontana argues that the decision has limited judicial precedence because of the highly contextual nature of the case. In particular, Fontana draws attention to both Google's decision to wholesale copy the API and the fact that Google is an extremely large software development company. By contrast, a smaller or open source developer's copying of functionality might have less market impact.[181] However, in critiquing the decision, Samuelson argues that the Federal Circuit had mistakenly revived the broad conceptualisation of copyright from *Apple v Franklin*. This reversal of precedent introduces new uncertainties into the enforcement of copyright law.[182] In addition, the *Oracle v Google* decision creates uncertainty for open source developers who assume APIs and other functional aspects of programming languages fall outside the realm of copyright protection.[183] Chapters Four, Five and Six of this thesis will more fully consider how open source licensors practically adapt to these requirements.

2.5.2 *Exceptions to Copyright Infringement for Software in the EU*

By contrast to the broad fair use exception under copyright law, the Software and Information Society Directives contain fair dealing exceptions. These exceptions are designed to achieve a balance between protecting original works of authorship in software and ensuring that copyright is not used for anti-competitive purposes.[184] In particular, the Software Directive

---

[177] *Oracle America, Inc. v Google Inc.* (Fed. Cir., No No. 17-1118, 2018) 34-6.

[178] In particular, Judge Alsup noted that Oracle and Sun had already acquired patent protection for parts of the Java API functionality

[179] *Oracle America, Inc. v Google Inc.* 872 F.Supp.2d 974 (Dist. Court, ND California 2012), 996; Pamela Samuelson, 'Copyrightability of Java APIs Revisited' (2015) 58(3) *Communications of the ACM* 22–24 24.

[180] *Google, Inc. v Oracle America, Inc.: Supreme Court* 135 S. Ct. 2887 (2015).

[181] Richard Fontana, above n141, 60.

[182] Pamela Samuelson, 'Functionality and Expression in Computer Programs: Refining the Tests for Software Copyright Infringement' (2016) 31(2) *Berkeley Technology Law Journal* 1215–1300 1222, 1299.

[183] Clark D. Asay, 'Software's Copyright Anticommons' (2016) 66(2) *Emory Law Journal* 265–332 314.

[184] Thorsten Käseberg, *Intellectual Property, Antitrust and Cumulative Innovation in the EU and the US* (Bloomsbury Publishing, 2012) 17, 126.

includes limited exceptions for reverse engineering copyrighted software (as opposed to a general fair dealing exemption). Further, the Software Directive prohibits licensing that would prevent either reverse engineering or the disclosure of underlying interoperability information for software.[185]. Finally, the Software Directive also reserves the right to allow for courts[186] to rely on antitrust provisions to mandate the release of interface details to encourage interoperability.[187] Further, the Information Society Directive also supplies a series of exceptions. In particular, Article 5(1) of the Information Society Directive creates a general exception for temporary acts of reproduction as part of a technological process.[188] This exception covers the copying of software into memory, or the transmission of software over a network, provided that the transmission is lawful. Of additional relevance to this thesis is the exception for the right of reproduction and communication provided by Article 5(3) for scientific research.[189]

The impact of these exceptions on open source software development is unclear. However, in some circumstances reverse engineering of software functionality may be impossible without directly copying source code. Accordingly, standardisation through licensing is required to achieve certain functionality.[190] An important future development of EU transnational copyright law is the passage of a new Copyright Directive. This new Copyright Directive was introduced as a component of the EU's Digital Single Market strategy to unify digital marketing, E-commerce and telecommunications policy.[191] One of the stated purposes of the new Copyright Directive is to assist the free flow of data throughout the EU. Further, the new Copyright Directive seeks to amend both the Information Society Directive and the Database Directive (discussed below) to support text and data mining (TDM), particularly for scientific research.[192] This reform was adopted due to the inflexibility of the current

---

[185] Directive 209/24/EC of the European Parliament and of the Council of 23 April 2009, Articles 5 and 6 & 9(1) respectively

[186] As well as the European Commission.

[187] *Treaty on the Functioning of the European Union*, opened for signature 7th February 1992, [2009] OJC 115/199 (entered into force 1st November 1993) Articles 102 and 103; *Microsoft Corp v Commission of the European Communities (T-201/04)* [2007] ECR II–3601.

[188] *Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights of the Information Society* [2001] *OJ L* 167/10 Article 5(1).

[189] *Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights of the Information Society* [2001] *OJ L* 167/10 Article 5(3)(a).

[190] Wolfgang Kerber and Heike Schweitzer, 'Interoperability in the Digital Economy' (2017) 8(1) *Journal of Intellectual Property, Information Technology and E-Commerce Law* 39–58 58.

[191] European Commission, *Communication from The Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* <https://eur-lex.europea.eu/legal-content/EN/TXT/?uri=celex%3A52015DC0192>.

[192] *Directive of the European Parliament and of the Council on Copyright in the Digital Single Market* [2016] *OJ C* 125 Recital 5.

exceptions as a means for researchers using TDM to avoid liability for legitimate uses. In particular, large scale TDM may trigger copyright infringement, and the limitation of existing exceptions to temporary reproductions would undermine the accuracy of data analysis.[193] However, there are concerns that many of the provisions of the Copyright Directive may act as an impediment to the stated goals of the Digital Single Market strategy. Specifically, Articles 3 and 4 of the new Copyright Directive creates an exception for the use of text and data mining software. This exception applies to both copyright and the *sui generis* database right. Whilst this optional requirement to implement this exception was mandated in the final version of the text, the structure of the exception itself also raises potential concerns. Further, academic commentators have noted the exception as written provides researchers with extremely limited scope to conduct TDM. First, the exception is still limited by the three step requirements imposed under international law. Secondly, and perhaps more importantly, the operation of the exception is limited to non-commercial scientific research only. This limitation is in spite of commercial scientific research facing the same technical limitations as non-commercial scientific research.[194] Subsequent chapters will discuss this effect in greater detail.

2.5.3 *Exceptions to Copyright Infringement in Australia and New Zealand*

In contrast with the US, but in a similar fashion to the EU, Australian copyright legislation provides an explicit list of fair dealing exemptions as opposed to a broad fair use exemption. These exemptions were introduced into Part 4A of the *Copyright Act 1968* as part of the *Digital Agenda Act*, and include two categories of exemptions:

1. those relating to incidental copying of software (such as using software and making backup copies);
2. those relating to reverse engineering and decompilation.[195]

Likewise, the New Zealand *Copyright Act 1994* was amended in 2008 to introduce specific exceptions for infringement with respect to copyrighted software for lawful use and decompilation.[196] However, Huang correctly notes that these exceptions are largely drafted with proprietary software in mind (as opposed to open source software).[197] In the limited case

---

[193] Christophe Geiger, Giancarlo Frosio and Oleksandr Bulayenko, 'Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?' (2018) 49(7) *IIC - International Review of Intellectual Property and Competition Law* 814–844 815.

[194] Benjamin Raue, 'Free Flow of Data? The Friction Between the Commission's European Data Economy Initiative and the Proposed Directive on Copyright in the Digital Single Market' (2018) 49(4) *IIC - International Review of Intellectual Property and Competition Law* 379–383 318-9.

[195] Anne Fitzgerald and Christina Cifuentes, above n73, 235.

[196] *Copyright Act 1994* (NZ) ss79-80, s80A, s80B, s80C, s80D, s81, s81A.

[197] Mary Huang, above n157, 54.

law that has been considered so far on software licensing in Australia, the primary focus has been on contract interpretation, as discussed within Section 2.4 above within the context of *Trumpet*.

2.6 CONCLUSION

Open source licensing has been advocated as a solution to ensure ongoing public access to open source software whilst protecting that software from being appropriated. This chapter demonstrates successful translation of open source licensing to software is heavily dependent on whether open source licences are enforceable. In particular, open source licences may be enforceable under either copyright or contract law. In addition, whether these licences operate in a consistent fashion across jurisdictions to reflect the cross jurisdictional nature of modern software development is an open question. Some case law from the US and Germany such as *Jacobsen v Katzer* have strongly supported open source licence enforceability under copyright and contract law. However, the divergent nature of precedent across both the US with cases, such as *MDY v Blizzard* means the enforceability of different licences remain uncertain. This uncertainty is particularly problematic with respect to the remedies that may be available for the breach of open source licences. In addition, although case law in Germany is supportive of open source licence enforcement, there is uncertainty as to whether this precedent will hold true across Europe.

In addition, there is extremely limited case law to support open source licence enforceability in Australia and New Zealand at present and the divergences in statutory protection for software may undermine the enforceability of these licences. Finally, the US fair use exemption has historically provided a useful 'safety valve' to provide developers with scope to reverse engineer software. This exception is significantly broader in scope than the equivalent fair dealing exceptions found in Europe, Australia and New Zealand. However, it is unclear whether this safety valve has been shut off in the aftermath of *Oracle v Google* and what impact this contraction will have on open source software development. Restrictions on reverse engineering and open source development may also be increased by the presence of patent rights, as discussed in the next chapter.

Chapter 3

PATENTS ON SOFTWARE AND METHODS: AT THE FRONTIERS OF PATENTABLE SUBJECT MATTER

3.1 INTRODUCTION

The subject of this chapter is the relationship between patents and software innovation and the impact (either theoretical or practical) that patents potentially have on software innovation. Chapter One alluded to the controversy flowing from early attempts to seek intellectual property protection for bioinformatics software and data. Chapter Two described the international and domestic frameworks for copyright protection as well as the disadvantages of copyright protection for software. This chapter addresses how each of these two factors drove the patenting of software. Whilst methods of operation were traditionally treated as falling outside the realm of patent protection the limitations of software copyright were a key driver of software developers seeking patent protection for computer programs. The discussion in this chapter informs the methodology described in Chapters Four, Five and Six. These chapters will examine how both patent and copyright law may be involved in the governance of open innovation in bioinformatics research.

This Chapter is organised into four sections. Section 3.2 provides an overview of patent law including the requirements for obtaining a patent and the exceptions to patent law as mandated by international law.[1] Section 3.3 supplies a comparative jurisprudential analysis of patent laws in the US, the EU, Australia and New Zealand. As for Chapter Two, this chapter adopts this comparative approach because the US was the first country to formally recognise patent protection for software. Further, until recently, the US has historically adopted an expansive interpretation of copyright law. By contrast, the EU (particularly the United Kingdom (UK)) has attempted to curtail software patenting through statutory intervention with varying degrees of success. Case law on software patents is slightly more developed in Australia and New Zealand than for open source licensing.

However, courts (and in the case of New Zealand, legislatures) have looked overseas for guidance on the boundaries of patentable subject matter. In addition, the deviation between different patent laws may have a significant impact on the effectiveness of open source licence provisions that control patent rights. Section 3.4 then provides an overview of the boundaries of patent protection and in particular, patent eligibility and the operation of research or experimental use exceptions. This analysis of patent eligibility will frame the subsequent empirical analysis in the second half of this thesis. Section 3.5 provides an overview of second tier or utility model patent systems, as well as *sui generis* regimes that have been implemented and proposed for software.

---

[1] *Marrakesh Agreement Establishing the World Trade Organization, annex IC, The Agreement on Trade Related Aspects of Intellectual Property Rights ('TRIPS Agreement')*, opened for signature 15th April 1994, 1867 UNTS 3 (entered into force 1st January 1995).

3.2 THE REQUIREMENTS FOR PATENT ELIGIBILITY

A patent is a limited monopoly granted to allow the patent holder to make use or exploit the subject matter of an invention, and to prevent others from exploiting that invention.[2] The requirements for patent protection are prescribed by Article 27.1 of the TRIPS Agreement. It mandates that patents are only available for inventions which are novel, inventive and have industrial applicability or utility. These requirements are normally questions of fact (with utility patent models in some jurisdictions lowering the requirement for patentability from inventiveness to innovativeness),[3]. However, there are three contentious requirements with respect to patent eligibility. These requirements are the extent of disclosure required in respect of the patent claims, patent eligibility and the exceptions to patent law. Each area of contention is addressed with reference to the relevant section in the TRIPS Agreement.

3.2.1 *Sufficiency of Disclosure*

Article 29 of the TRIPS Agreement mandates that signatories will ensure that a 'patent application shall disclose the invention in a manner sufficiently clear and complete for it to be carried out by a person skilled in the art'. It further provides that signatories *may* introduce 'best mode disclosure', whereby the applicant is required to indicate 'the best mode for carrying out the invention known to the inventor'.[4] The EU, the US, Australia and New Zealand have all arguably introduced the best mode requirement into national patent law.[5] However, the level of disclosure for certain forms of technology often presents unique challenges, particularly for genomic and software patents.[6] Although this chapter does not address the question of sufficiency of disclosure in any further detail, the next chapter addresses the question of disclosure with respect to the interaction between patent and open source software rights.

---

2 Thorsten Käseberg, *Intellectual Property, Antitrust and Cumulative Innovation in the EU and the US* (Bloomsbury Publishing, 2012) 15.

3 Philip Leith, *Software and Patents in Europe* (Cambridge University Press, 2011) 173.

4 Nefissa Chakroun, *Patents for development: improved patent information disclosure and access for incremental innovation* (Edward Elgar Publishing, 2016) 62.

5 *35 U.S. Code § 101 - Inventions patentable*; *Convention on the Grant of European Patents*, opened for signature 5th October 1973, 1065 UNTS 254 (entered into force 7th October 1977) article 83; *Patents Act 1990* (Cth) section 40(2); *Patents Act 2013* (NZ) section 39.

6 Viola Prifti, *The Breeder's Exception to Patent Rights: Analysis of Compliance with Article 30 of the TRIPS Agreement* (Springer, 2015) 47.

3.2.2 *The Scope of Patent Protection and Definition of Invention*

Perhaps the most contentious aspect of patent protection is determining the scope of patentable inventions. Article 27 mandates that signatories must offer protection on all forms of invention, including machines, substances, processes and products, that meet the above requirements for patentability. This definition itself (at least in common law jurisdictions such as the UK) evolved from the original English *Statute of Monopolies* which prohibited monopolies except for those 'letters patent' that disclosed a new 'manner of manufacture'.[7] Prior to this, the Italian city state of Venice had granted patents on useful inventions, although the time that the patent was granted for was variable. Further, the patent holder had to establish that the invention could not only work but that it was useful when reduced to practice.[8] However, over time the scope of patentable subject matter has expanded to encompass other forms of technology, including nanotechnology, biotechnology, software and diagnostic methods (all of which are of considerable significance to this thesis).[9]

3.2.3 *Exceptions to Patent Law*

Article 30 of the TRIPS Agreement provides signatory nations with the capacity to introduce exceptions to patent rights. Although Article 30 does not provide an inclusive list of exceptions, it provides that any exceptions must be subject to the following conditions:

1. The exception should not unreasonably conflict with a normal exploitation of the patent;
2. The exception should not unreasonably prejudice the legitimate interests of third parties;
3. The exception should take account of the legitimate interests of third parties.

As Sections 3.4.1, 3.4.2 and 3.4.3 of this chapter reveal, different jurisdictions have taken divergent approaches to the implementation of Article 30 exceptions. This divergence is particularly apparent in the context of exceptions for the use of patented technology for research related purposes (which are of perhaps the greatest importance to this thesis). As Chapter Four explores, the existence of experimental use exceptions may be treated as structural mechanisms to encourage the voluntary exchange of patented technologies.[10]

---

[7] *Statute of Monopolies 1623* (UK) 21 Jac s6; Edward C. Walterscheid, 'Early Evolution of the United States Patent Law: Antecedents (Part 2), The' (1994) 76(11) *J. Pat. & Trademark Off. Soc'y* 849 877.

[8] Mario Biagioli, 'Patent Republic: Representing Inventions, Constructing Rights and Authors' (2006) 73(4) *Social Research: An International Quarterly* 1129–1172 1133.

[9] Thorsten Käseberg, above n2 15.

[10] Nicole Ziegler, Oliver Gassmann and Sascha Friesike, 'Why do Firms Give Away their Patents for Free?' (2014) 37(2) *World Patent Information* 19–25 3.

Before discussing experimental use exceptions and scientific research, it is necessary to examine the divergence in patent laws between different jurisdictions. This deviation is important because differences in patent law, as well as the regional grant of patents, may influence how developers collaborate across jurisdictions.[11]

3.3 COMPARATIVE LAW ON THE SCOPE OF PATENT ELIGIBILITY

3.3.1 *The United States*

3.3.1.1 The Expansion of Patentable Subject Matter

In the US, patents are governed by the federal *Patents Act*, which is given effect by the US Constitution and empowers the USPTO to grant patents on inventions that meet the requisite standards of novelty, usefulness and inventiveness. Along with natural discoveries and laws of nature, abstract ideas traditionally fell entirely outside the scope of intellectual property protection, with the expression of ideas being traditionally considered the domain of copyright law. To distinguish between the two concepts, the US Supreme Court in the mid 19th century held that an abstract idea, apart from its implementation, is not patent-eligible.[12] This test was formalised in *Gottschalk v Benson* as the 'machine or transformation' test.[13] Initially US courts were unwilling to grant patents for software, arguing that the National Commission on New Uses of Copyrighted Works (CONTU), as discussed in Chapter Two, had established that the appropriate regime for protecting software was copyright law.

In *Gottschalk v Benson*,[14] the Supreme Court rejected a patent on an algorithm for the conversion of digital numbers into binary digits for mathematical calculations on the grounds that mathematical algorithms were part of the 'liberal arts' and not the 'useful arts'. The Court expressed the belief that because algorithms were implementations of mathematical formulae, they were to be treated as patent ineligible subject matter.[15] It took issue with the claim structure of the patent in question on the grounds that it pre-empted both known and unknown uses of the algorithm with a 'general purpose digital computer'.[16] Pamela Samuelson primarily critiques the Supreme Court's decision in *Gottschalk* as lacking a consistent judicial framework. Nevertheless, Samuelson notes that this pre-emption argument amounted to an attempt by the

---

[11] James Ernstmeyer, 'Does Strict Territoriality Toll the End of Software Patents Note' (2009) 89(4) *Boston University Law Review* 1267–1304 1275.

[12] *Le Roy v Tatham* 55 US (14 How.) 156 (1852); *O'Reilly v Morse* US (15 How.) 62 (1854).

[13] *Gottschalk v Benson* 409 U.S. 63 (1972), 70.

[14] *Gottschalk v Benson* 409 U.S. 63 (1972), 67.

[15] *Gottschalk v Benson* 409 U.S. 63 (1972), 69-71.

[16] *Gottschalk v Benson* 409 U.S. 63 (1972), 67.

Supreme Court to justify prohibiting software patents.[17]

In the subsequent case of *Parker v Flook*,[18] the Supreme Court again invalidated an algorithm patent, this time directed to the implementation of an alarm clock for calculating catalytic conversions of hydrocarbons. It held that the subject matter for the novel part of the invention claimed was merely the implementation of the algorithm.[19] However, unlike the Court in *Gottschalk*, the Supreme Court noted that software patents were 'not undesirable as a matter of policy'. In addition, the Supreme Court noted the simultaneous debate about the limitations of scope of functional copyright protection for software as discussed in Chapter Two. The Supreme Court therefore left open the possibility of software patents being granted.[20] Subsequently in *Diamond v Diehr*, the Supreme Court held that a similar algorithm to that considered in *Parker* (this time for curing rubber) was valid patentable subject matter. It distinguished between the two cases on the grounds that despite the fact the algorithm was in the public domain,[21] the implementation of the algorithm was patent eligible.[22] In part, this change in the patent eligibility rules for software can be attributed to the changes by the US Patent and Trademarks Office (USPTO)'s policy and practice in issuing software patent inventions. This policy change was driven by the desire to protect software inventions due to weakened copyright protection.[23]

The decision in *Diehr* followed the decision in *Diamond v Chakrabarty*[24]. In *Chakrabarty* the Supreme Court upheld a patent on a genetically engineered bacterium on the grounds that 'everything under the sun' (figuratively, any process or product) was patentable subject matter. Both of these cases significantly broadened the scope of patent eligibility to include inventions where there was some physical effect. After a trio of contemporaneous decisions from the Court of Customs and Patent Appeals which applied the *Diamond v Diehr* reasoning, this test became known as the *Freeman-Walter-Abele* test. This test was also known as the *machine or transformation* test (given that the test tied patent eligibility to the physical effect flowing from the patented invention).[25] Hazel Moir argues that applying this machine or transformation test,

---

[17] Pamela Samuelson, 'Benson Revisited: The Case Against Patent Protection For Algorithms and Other Computer Program-Related Inventions' (1990) 39(4) *Emory Law Journal* 1025–1154 1056.

[18] *Parker v Flook* 437 U.S. 584 (1978).

[19] *Parker v Flook* 437 U.S. 584 (1978), 595.

[20] *Parker v Flook* 437 U.S. 584 (1978), 595.

[21] The equation in question was the Arrhenius equation, which is used for calculating chemical reaction rates

[22] *Diamond v Diehr* 450 U.S. 175 (1981), 188.

[23] As described in Sections 2.2.2 and 2.2.3 of Chapter Two (Pamela Samuelson, above n17, 1093).

[24] *Diamond v Chakrabarty* 447 U.S. 303 (1980).

[25] *In re Freeman* 573 F.2d 1237 (C.C.P.A. 1978); *In re Walter* 618 F.2d 758 (C.C.P.A. 1980); *In re Abele* 684 F.2d 902 (C.C.P.A. 1982).

embedded software (or software tied to a particular hardware platform) would fall inside the realm of patent eligibility. By contrast, software running on a general purpose computer (where there is no specific connection between the effect of the software on the hardware) would not.[26] This interpretation is supported by Samuelson. In particular, Samuelson notes that the test from *Diehr* limited the expansion of the scope of patentable subject matter to traditionally patentable industrial processes including a computer program.[27]

The changing nature of the Supreme Court's approach can also be attributed to the declining prominence of Justice Stevens, who wrote the majority judgments in *Gottschalk* and *Flook*. Due to the changing composition of the Court in the years between *Gottschalk* and *Benson*, Justice Stevens wrote the minority judgment in *Diamond v Diehr*. In his dissenting judgment, Justice Stevens argued that drafting was the key distinction between the three algorithms under consideration. Justice Stevens further noted that in *Diehr* was that the applicant characterised the use of the algorithm to calculate rates of rubber curing as the novel invention. By contrast, the algorithm itself was not novel, but instead involved the use of the two public domain algorithms together in a fashion that was novel.[28] However, Justice Stevens held that just because the two components were combined in a way that was novel, that combination did not render the invention as a whole patent eligible subject matter.[29] Despite the fact that Justice Stevens was in the minority in *Diamond v Diehr*, his decision had a significant impact on shaping later authority regarding patentable subject matter.[30] This effect is addressed later in this section.

Contemporaneous with the *Diehr* and *Chakrabarty* decisions was the establishment of the Federal Circuit Court of Appeals, a specialist tribunal that was resourced to hear patent appeals from state courts. The Federal Circuit was established from the Court of Customs and Patent Appeals to provide consistency in the hearing of patent appeals from state courts.[31] However, the Federal Circuit also operates on the presumption that all patents granted by the US Patents and Trademarks Office (USPTO) are prima facie valid, resulting in a higher number of patents being upheld by the Federal Circuit when challenged.[32] Finally, the USPTO

---

[26] Hazel V. J. Moir, *Do Patent Systems Improve Economic Well-Being? An Exploration of the Inventiveness of Business Method Patents* (PhD Thesis, The Australian National University, 2008).

[27] Pamela Samuelson, above n17, 1029.

[28] *Diamond v Diehr* 450 U.S. 175 (1981), 209.

[29] *Diamond v Diehr* 450 U.S. 175 (1981), 218-9.

[30] Timothy R. Holbrook and Mark D. Janis, 'Patent-Eligible Processes: An Audience Perspective' (2014) 17(2) *Vanderbilt Journal of Entertainment and Technology Law* 349–384 365.

[31] Rochelle Cooper Dreyfuss, 'The Federal Circuit: A Case Study in Specialized Courts' (1989) 64(1) *New York University Law Review* 1 21.

[32] *35 U.S. Code § 282 - Presumption of validity; defenses*; Doug Lichtman & Mark A. Lemley, 'Rethinking Patent Law's Presumption of Validity' (2007) 60(1) *Stanford Law Review* 45 69.

itself received additional funding and infrastructure in the 1980s to handle the influx of software patent applications.[33] In concert, these factors meant that developers were significantly more likely to have their patent applications granted and upheld, provided that they could overcome the physical effect requirement from *Diehr*.

In *In re Alappat*,[34] a majority of the Federal Circuit dispensed with this requirement. The Federal Circuit held that the appropriate test for patent eligibility was instead whether the patented software produced a 'useful, concrete or tangible result', even on a generic computer. The Federal Circuit went beyond the reformulated test from *In re Alappat* in the subsequent case of *State Street v Signature Financial Group* ('*State Street*'),[35]. Instead, the Federal Circuit held that there was no special test for patent eligibility for software provided that the software package produces a useful result. *In re Alappat* and *State Street* dramatically expanded the scope of protection for computer programs, including financial decision making and data management methods. These decision making patents have subsequently become known as business method patents within patent literature. *In re Alappat* and *State Street* also opened the scope of patent eligibility for software running on general purpose computers.[36]

3.3.1.2 The Contraction of Patentable Subject Matter

However, the past decade has witnessed a gradual rejection of the expansion of patentable subject matter, beginning with the Supreme Court's decision in *Laboratory Corporation of America Holdings v Metabolite Labs Incorporated* ('*Laboratory Corporation v Metabolite Labs*').[37] In considering whether a vitamin deficiency diagnostic test patent was invalid for attempting to claim a 'monopoly over a basic scientific relationship', Justice Stevens (once again in dissent) held that the Supreme Court had never determined on whether 'a useful or tangible result' was the appropriate test for patent eligibility.[38] A majority of the court held that there was no risk of related diagnostic tests being impeded by the patent in question. Nevertheless, Mark Lemley and others argued that Justice Stevens' dissent sowed the seeds for the Supreme Court's contraction of patentable subject matter, beginning with *Bilski v Kappos* ('*Bilski*'). *Bilski* was a series of cases heard in both the Federal Circuits and Supreme

---

[33] Noam Shemtov, *Beyond the Code: Protection of Non-Textual Features of Software* (Oxford University Press, 2017) 165-6.

[34] *In re Alappat* 33 F.3d 1526 (Fed. Cir. 1994), 1545.

[35] *State Street Bank Trust Co. v Signature Financial Group Inc.* 149 F. 3d 1368 (Fed. Cir. 1998).

[36] Benjamin J. McEniery, 'An Empirical Study of Business Method Patent Applications Filed in Australia 2000-2009' (2012) -(89) *Intellectual Property Forum*; Hazel V. J. Moir, above n26, 138-178 <https://openresearch-repository.anu.edu.au/handle/1885/49313>.

[37] *Laboratory Corporation of America Holdings v Metabolite Labs Incorporated* 548 U.S. 124 (2006).

[38] *Laboratory Corporation of America Holdings v Metabolite Labs Incorporated* 548 U.S. 124 (2006), 137.

Courts. As with *State Street Bank*, *Bilski* concerned the patentability of a computer implemented financial trading system. In contrast to *State Street Bank*, the Federal Circuit held that the 'useful or tangible result' test was too broad to convey the scope of patent eligibility. Instead, the Federal Circuit held that patent eligibility rests with whether the invention is 'tied to a particular thing or apparatus' or 'transforms an article to a different state or thing'.[39] Dennis Crouch and Robert Merges argue that in reaching this conclusion, the Federal Circuit had resurrected the 'machine or transformation' test from *Diamond*. Accordingly, that a ruling from the Supreme Court would be required to resolve the conflicting jurisprudence.[40]

The Supreme Court in the *Bilski* appeal followed the Federal Circuit in significantly deviating from the jurisprudence established in *State Street Bank*. However, clear judgment on the scope of patent eligibility did not emerge from the Court's decision.[41] Although Justice Breyer, delivering the judgment of the majority, acknowledged that allowing the patenting of abstract principles would 'wholly pre-empt' the public's access to basic tools of scientific research. His Honour nevertheless fell short of endorsing the 'machine or transformation' test. Instead, Justice Breyer characterised the 'machine or transformation' test as an important example of what might be indicative of patent eligibility, but not an exclusive test. Characterising it as exclusive would expand the explicit statutory limitations on laws of nature, physical phenomena and abstract ideas.[42] In determining what features would be indicative of patentability, the majority directly contrasted the patent in question in *Benson* with the patent in question in *Diehr*. In the case of the former, the majority defined the former as a pure algorithm which therefore fell outside the scope of patentable subject matter. By contrast, the latter was the algorithm had a direct change in the physical state of rubber. The majority agreed that purely abstract subject matter was not patentable, whilst still retaining support for patents on 'software' and 'advanced diagnostic techniques'.[43]

In contrast, Justice Stevens, true to his earlier judgments, advocated a return to the 'machine or transformation' test as the sole test for patent eligibility. With respect to the plurality judgment, Justice Stevens argued that to characterise the presence of a physical effect as an indicator of patent eligibility would lead to an absurd result that any process that was 'non-obvious, novel and described with specificity' would be patentable.[44] Justice Stevens

---

[39] *In re Bilski* 545 F.3d 943 (2009), 954, 959.

[40] Dennis Crouch and Robert P. Merges, 'Operating Efficiently Post-Bilski by Ordering Patent Doctrine Decision-Making' (2010) 25(4) *Berkeley Technology Law Journal* 1673–1692 1691.

[41] *Bilski v Kappos* 130 S. Ct. 3218 (2010), 3231, 3259.

[42] *Gottschalk v Benson* 409 U.S. 63 (1972), 70; *Parker v Flook* 437 U.S. 584 (1978), 588-9; *Bilski v Kappos* 130 S. Ct. 3218 (2010), 3226.

[43] *Bilski v Kappos* 130 S. Ct. 3218 (2010), 3237 per Justice Kennedy.

[44] *Bilski v Kappos* 130 S. Ct. 3218 (2010), 3235 per Justice Stevens.

advocated returning patentable subject matter to its historical roots and excluding 'methods of doing business' from the scope of patent protection under section 101 of the US Patent Act.[45] Both the plurality and minority decisions of the Supreme Court in *Bilski* represented a significant shift away from the broad test for patent eligibility from *State Street Bank*.[46] This shift was observable in the USPTO guidelines released after *Bilski*. These guidelines noted that although the 'machine or transformation' test was only one factor indicating patent eligibility, no court had ever ruled 'that a method claim... that lacked a machine or transformation was patent eligible'.[47] A 5-4 majority of the Supreme Court still explicitly endorsed software as patent eligible. However, the subsequent decision of the Supreme Court in *Mayo v Prometheus* ('*Mayo*')[48] further eroded this boundary.

In *Mayo*, the Supreme Court further reinforced the comparison between *Flook* and *Diehr* to distinguish between patentable and non-patentable inventions using a machine or transformation test. *Mayo* concerned the patent eligibility of a natural phenomenon specifically focusing on a diagnostic test for optimising drug doses for patients with irritable bowel syndrome. At first and second instance and considering the majority judgment in *Bilski*, the Federal Circuit held that the patent in question satisfied the requirements for the 'machine or transformation' test. The diagnostic test was arguably directed towards a naturally occurring effect (that is, detecting 'irritable bowel syndrome'). However, its application in the invention towards a specific diagnostic purpose was sufficient to avoid exclusion under section 101. In supporting the invention as patentable subject matter, the Federal Circuit followed the judgment of the Supreme Court in *Bilski* in refusing to use the 'machine or transformation' test as a way of excluding diagnostic methods from the scope of patentability.[49] However, on appeal, the Supreme Court invalidated the patent, holding that the interaction between the drug metabolite levels and side effects was a recognised effect within existing medical literature. Therefore, the claimed invention did not amount to an 'inventive concept' that would otherwise extend the natural principles at play.[50]

Timo Minssen and Robert Schwartz argue that in rejecting the patent, the Supreme Court went further on the question of patentable subject matter than it did in *Bilski*. Minssen and Schwartz note that the claims under examination in *Mayo* were representative of standard

---

[45] *35 U.S. Code § 101 - Inventions patentable*.

[46] Anton Hughes, 'Bilski v Kappos Case Brief' (2010) 20(1) *Journal of Law, Information and Science* 206–213.

[47] United States Patent and Trademark Office, 'Interim Guidance for Determining Subject Matter Eligibility for Process Claims in View of Bilski v. Kappos' (27 July 2010) 75(43922) *F.R.*

[48] *Mayo Collective Services v Prometheus Laboratories* 132 S. Ct. 1289 (2012), 1298-9.

[49] *Prometheus Laboratories v Mayo Collaborative Ser.* 581 F.3d 1336 (2009), 1345-7; *Prometheus Laboratories v Mayo Collaborative Ser.* 628 F.3d 1347 (2010), 1355; *Bilski v Kappos* 130 S. Ct. 3218 (2010), 3237 per Justice Kennedy.

[50] *Mayo Collective Services v Prometheus Laboratories* 132 S. Ct. 1289 (2012).

diagnostic and personalised medicine claims that had previously been issued by the USPTO.[51] The controversy from *Mayo* was exacerbated by the fact that the Supreme Court did not issue a clear and definitive test as to when the application of a natural phenomenon would amount to a patentable invention. In considering the machine or transformation test as a factor in patent eligibility, the Supreme Court described *Benson* and *Flook* as patent claims that merely recited natural laws. Whilst *Diehr* involved an algorithm as well, it also contained an application of that algorithm.[52] In *Mayo*, the Supreme Court relied upon this dichotomy to distinguish between therapeutics, which involved an application of natural laws. By contrast, diagnostics such as the claimed invention under consideration only involved measuring biological responses to medical intervention and were not patentable subject matter.[53]

However, as Eisenberg notes, the Supreme Court missed the opportunity in *Mayo* to develop a test to ascertain when diagnostic processes would be patentable beyond a nebulous distinction between therapeutics and diagnostic methods.[54] Further, Eisenberg argues that the decision in *Mayo* left considerable confusion when considering the ineligibility of abstract process patents such as software patents as opposed to natural phenomena.[55] As discussed in Chapter Four, this decision created significant complications for determining the eligibility of bioinformatics patents at the USPTO. It was not until *Alice Corporation v CLS Bank* (‘*Alice*’) that the Supreme Court sought a solution to the impasse created over whether abstract ideas were patent eligible.[56] Unlike *Bilski*, the en banc (unanimous) decision of the Supreme Court in *Alice* represented an absolute repudiation of patents that apply abstract processes using a generic computer. The Supreme Court’s decision also severely limited the scope of business method patents such as the financial transaction patent that was in issue in this case. In so holding, the Court expressed concern regarding the pre-emption of generic computational techniques. The Court also expressed concern regarding the difficulty of delineating the distinction between patents that ‘claim the building blocks of human ingenuity and those that integrate the blocks into something more’.[57] Accordingly, the Supreme Court developed a two part test to determine patentable subject matter. The first step of this test involves determining

---

[51] Timo Minssen and Robert M. Schwartz, ‘US Patent Eligibility in the Wake of Bilski v Kappos: “Business as Usual” in an Age of New Technologies?’ (2011) 30(1) *Biotechnology Law Report* 3–56 35.

[52] *Mayo Collective Services v Prometheus Laboratories* 132 S. Ct. 1289 (2012), 1294.

[53] *Mayo Collective Services v Prometheus Laboratories* 132 S. Ct. 1289 (2012), 1302.

[54] Rebecca S. Eisenberg, ‘Prometheus Rebound: Diagnostics, Nature, and Mathematical Algorithms’ (2013) 122(1) *Yale Law Journal Online* 341–837 343; Rebecca S. Eisenberg, ‘Diagnostics Need Not Apply’ (2015) 21(1) *Boston University Journal of Science and Technology Law* 256–286 278.

[55] Rebecca S. Eisenberg, above n54 278.

[56] *Alice Corporation v CLS Bank* 573 U.S. ¸134 2347 (2014).

[57] *Alice Corporation v CLS Bank* 573 U.S. ¸134 2347 (2014), 2354.

whether the patent in question is directed to a 'patent ineligible abstract idea'.[58] If this question is answered in the affirmative, the second part of the test requires the court to consider whether enough has been added to transform the abstract claim into patent eligible subject matter.[59]

The Supreme Court held that performing 'well-understood, routine, conventional activities previously known to the industry', such as those described in the appellant's patents, were not patent eligible when performed with a generic computer.[60] As discussed previous, a large number of business method patents were entirely invalidated by the decision in *Alice*. In addition, the decision in *Alice* also led to an increase in the non-renewal and invalidation of software patents.[61] However, there is more recent Federal Circuit authority which suggests that software related inventions may be capable of passing the two part test from *Alice*.[62] Of importance to this thesis is the way in which the decisions in *Alice* and *Mayo* have been used by the Federal Circuit in concert to deny a broader class of diagnostic patents. These decisions have been combined with the precedent from *Myriad Genetics v Association for Molecular Pathology*.[63] In *Myriad Genetics v AMP*, the Supreme Court applied the same natural principle test to deny patents on naturally occurring DNA sequences. The combined effect of these decisions was demonstrated in *Ariosa Diagnostics v Sequenom* ('*Ariosa*').[64] *Ariosa* concerned a patent directed towards non-invasive prenatal testing, where a blood test from a pregnant woman could be used to determine whether her child would suffer from particular genetic defects. The Federal Circuit denied protection for the patent on the grounds that the subject matter contravened both the prohibition on patents directed towards natural phenomena (with respect to DNA) and abstract processes (with respect to the diagnostic method).[65] However, as both Eisenberg and Dreyfuss note, the failure to grant a patent for the method in question was heavily disputed amongst the scientific community. In particular, because of the inventiveness involved in replacing amniotic fluid testing (which carries the risk of miscarriage) with a blood test, Eisenberg and Dreyfuss argue the patent should have

---

[58] *Alice Corporation v CLS Bank* 573 U.S. ¸134 2347 (2014), 2352.

[59] *Alice Corporation v CLS Bank* 573 U.S. ¸134 2347 (2014), 2354.

[60] *Alice Corporation v CLS Bank* 573 U.S. ¸134 2347 (2014), 2359.

[61] James E. Daily, 'Alice's Aftermath: Changes in Patentee Behavior since Alice v. CLS Bank Bridging the Gap between the Federal Courts and the United States Patent & Trademark Office: The Journal of Science and Technology Law Symposium' (2017) 23() *Boston University Journal of Science and Technology Law* 284–303 301.

[62] *DDR Holdings LLC v Hotels.com L.P.* 773 F.3d 1245 (Fed. Cir. 2014); *Enfish, LLC v Microsoft Corp* 822 F.3d. 1327 (Fed. Cir. 2016).

[63] *Association for Molecular Pathology v Myriad Genetics, Inc.* 133 S. Ct. 2107 (2013).

[64] *Ariosa Diagnostics, Inc. v Sequenom, Inc.* 788 F.3d 1371 (Fed. Cir. 2014); *Sequenom, Inc. v Ariosa Diagnostics, Inc.* 136 S. Ct. 2511 (2016).

[65] *Ariosa Diagnostics, Inc. v Sequenom, Inc.* 788 F.3d 1371 (Fed. Cir. 2014), 1373.

passed the threshold for patentable subject matter.[66] Ultimately, by attempting to restrict the scope of patent protection for intangible methods, it appears that the Supreme Court has created greater uncertainty into the test for patentable subject matter.

3.3.2 *Patentable Subject Matter in Europe*

The European Patent Convention (EPC) defines the boundaries of patent eligibility in Europe and establishes the jurisdiction of the European Patent Office (EPO) and its Technical Boards of Appeal.[67] This treaty creates a bifurcated patent system for EPC member states that is separate from the domestic patent system. In other words, the EPC does not create 'European wide' patents. Instead, the EPC provides a mechanism for the uniform grant and maintenance of a particular patent. Further, the EPC mandates that member states introduce uniform rules for patent eligibility.[68] Prior to the establishment of the EPC, there were no express limitations on the technological boundaries of patentable subject matter. For example, in the UK, the *Patents Act 1949* (UK) mandated that patentable inventions included any invention related to a 'manner of new manufacture' (which was the English common law test for patentable subject matter).[69] Although this 'manner of manufacture' does not explicitly relate to software, Phillip Leith notes that there were attempts to argue that a novel algorithm (as opposed to a novel computer system) could fulfil this manner of manufacture requirement.[70]

Likewise, German patent law ('*Patentgesetz*') explicitly protected products and processes that were new, useful and capable of industrial application, or contain an inventive step, via its second tier utility patent system.[71] However, the establishment of the EPC in 1973 required signatories to introduce a unified standard for patentable subject matter.[72] Crucially, Article 52 requires member states to allow for patent protection for all fields of technology that are new, useful and capable of industrial application. Further, Article 52 also excludes certain technological fields from protection as inventions. These exclusions include computer programs 'as such' and methods of medical treatment. Plant and animal varieties (with the

---

[66] Rebecca S. Eisenberg, above n54; Rochelle Cooper Dreyfuss, 'Reconsidering Experimental Use' (2016) 50(4) *Akron Law Review* 699–724 711.

[67] It should be noted that the European Patent Convention is not a part of European Union law. For example, Switzerland, Liechtenstein, Turkey, Monaco, Iceland, Norway, Albania and Serbia are members of the EPC but not of the EU.

[68] Hanns Ullrich, 'Patent Protection in Europe: Integrating Europe into the Community or the Community into Europe?' (2002) 8(4) *European Law Journal* 433–491 436.

[69] *Patents Act 1949* (UK) section 101.

[70] Philip Leith, above n3, 11-3.

[71] *Patentgesetz [German Patent Act] 1980* section 1; *Gebrauchsmustergesetz [Utility Model Act] 2009* section 1.

[72] *Convention on the Grant of European Patents*, opened for signature 5th October 1973, 1065 UNTS 254 (entered into force 7th October 1977) Article 52.

exception of microbiological organisms as per the European Biotechnological Inventions Directive[73]) are also excluded as patentable subject matter.[74] Despite the initially strict interpretation of the Article 52 and 53 exclusions, the various Patent Technical Boards of Appeal responsible for software and diagnostic methods have sought to expand the scope of patentability.[75]

The Technical Boards of Appeal assigned to review software patents, Board 3.5.1, first had the opportunity to consider an appeal on patent eligibility in *Vicom*. This patent concerned an algorithm for improving the quality of digital images on a display.[76] The patent examiner at the EPO denied the patent on the grounds that it was directed to a mathematical algorithm, and therefore fell outside the realm of patentability. However, Board 3.5.1 granted the patent on the grounds that the improvement in image quality amounted to a technical result directed by a technical process.[77] In other words, although algorithms alone fell outside the realm of patentability, technical processes relying on these algorithms were patentable. The Technical Boards of Appeal arguably reached the correct conclusion in extending patentability to an invention with a physical effect. However, Sigrid Sterckx and Julian Cockbain, as well as Leith, have described the reasoning of the Board of Appeal as suspect. In particular, if an applicant drafted an algorithm patent as having a technical result pursuant to the rule in *Vicom*, the prohibitions on patenting under Articles 52 and 53 could be completely bypassed.[78] As Justine Pila notes, the drafters of the EPC had expressly *excluded* technical character as a test for determining patentable subject matter.[79] From this decision, there began to emerge a split in the authority emerging from the Technical Board of Appeals, which first became apparent in the subsequent *IBM* decision.[80] Acknowledging that Articles 52 and 53 did not forbid mixed media patent claims, the Technical Board of Appeals in *IBM* held that it could not subvert the nature of these exclusions entirely.[81] Instead, the Technical Board of Appeals held

---

[73] *Directive 98/44/EC of the European Parliament and of the Council of 6 July 1998 on the Legal Protection of Biotechnological Inventions* [1998] *OJ L* 213/13.

[74] *Convention on the Grant of European Patents*, opened for signature 5th October 1973, 1065 UNTS 254 (entered into force 7th October 1977) Articles 52(2)(c), 53(a)-(c).

[75] Philip Leith, 'Patenting Programs as Machines' (2007) 4(2) *SCRIPTed: A Journal of Law, Technology and Society* 214–226 220; Sigrid Sterckx and Julian Cockbain, 'The Patentability of Computer Programs in Europe: An Improved Interpretation of Articles 52(2) and (3) of the European Patent Convention' (2010) 13(3) *The Journal of World Intellectual Property* 366–402 368.

[76] *T-208/84* (15th August 1986) [1987] O.J. EPO ('Vicom/Computer-related invention').

[77] *T-208/84* (15th August 1986) [1987] O.J. EPO ('Vicom/Computer-related invention') 26.

[78] Sigrid Sterckx and Julian Cockbain, above n75; Philip Leith, above n3, 31.

[79] Justine Pila, 'Dispute Over the Meaning of 'Invention' in Article 52(2) EPC: The patentability of computer-implemented inventions in Europe' (2005) 36(2) *IIC: International Review of Industrial Property & Copyright Law* 171 186.

[80] *T-38/86* (14th February 1989) [1990] OJ EPO ('IBM/Text Processing').

[81] *T-38/86* (14th February 1989) [1990] OJ EPO ('IBM/Text Processing') 25.

that an algorithm would only be patentable where the invention made a *contribution in the art to a field not excluded from patentability*.[82]

The contribution approach was similar to the machine or transformation test, and provided a more consistent means of filtering software patents than the technical result approach. However, the passage of the TRIPS Agreement in 1994 meant that signatories had to ensure that patents were available for all fields of technology.[83] This requirement resulted in a redrafting of Article 52 in 2000 to reflect the fact that EPC signatories were also signatories to the TRIPS Agreement. In addition, the *IBM II* series of cases amounted to (what was at the time) an alignment with US and Japanese patenting practice regarding computer programs.[84] In both of these cases, the Technical Board of Appeals 3.5.1 held that the EPO and its subsidiary bodies were not bound by the TRIPS Agreement in interpreting the scope of software patent law[85]. Nevertheless, Article 27 created friction in the interpretation of Articles 52 and 53.[86] The Technical Board of Appeals in both IBM cases held that software was patentable, provided that it produced some technical effect (even with a general purpose machine). This conclusion was reinforced by the subsequent decision of *Ricoh*,[87] which held that whether an effect was technical or non-technical would be determined by reference to the prior art.

Despite the line of EPO authority from *IBM* and *Ricoh*, there has also been a considerable divergence at the domestic level on the proper interpretation of the Article 52 and 53 exclusions in the UK. This fragmentation is particularly noticeable in the UK and Germany. Although the Court of Appeal and the Patent Court had at various stages rejected the 'technical effect' test before the passing of the TRIPS Agreement,[88] it was Lord Justice Jacob of the Court of Appeal who led a strong critique of this approach in *Aerotel/Macrossan*.[89] Both of these cases were heard together as they referred to software utilising well-known hardware to communicate and procure documents and both were also to be referred to the EPO Enlarged Board of Appeal.

---

[82] *T-38/86* (14th February 1989) [1990] OJ EPO ('IBM/Text Processing') 12.

[83] Sigrid Sterckx and Julian Cockbain, above n75, 381; Justine Pila, 'Software Parents, Separation of Powers, and Failed Syllogisms: A Cornucopia from the Enlarged Board of Appeal of the European Office' (2011) 70(2) *Cambridge Law Journal* 203–228 204-5.

[84] *T-1173/97* (1st July 1998) [1997] OJ EPO ('Computer program product/IBM'); *T-935/97* (4th February 1999) [1997] EPOR ('Computer program product II/IBM').

[85] *T-1173/97* (1st July 1998) [1997] OJ EPO ('Computer program product/IBM') 6.4-6.5; *T-935/97* (4th February 1999) [1997] EPOR ('Computer program product II/IBM') 2.4.

[86] Sigrid Sterckx and Julian Cockbain, above n75, 374.

[87] *T-172/03* (27th November 2003) [2003] 'RICOH/Order Management') 8-10.

[88] *Merrill Lynch Inc.'s Application* [1988] RPC 1; *Gale's Application* [1991] RPC 305; *Fujistu Limited's Application* [1997] RPC 608.

[89] *Aerotel Ltd v Telco Holdings Ltd Ors* [2006] EWCA Civ 1371; *Macrossan v Comptroller General of Patents, Designs and Trade Marks 2005* [2006] EWHC 705 (Ch).

Instead of the 'technical effect' approach, Lord Justice Jacob enunciated a four step test. This test was in turn an adaptation of the four step test used by the UK Intellectual Property Office (IPO) for assessing the patent eligibility of an invention during examination:[90]

1. properly construe the claim and determine what the monopoly is;
2. identify the actual contribution by examining the effect on the known art and what the inventor added to human knowledge by looking to the substance not form of the claim;
3. ask whether the contribution is solely within or consists solely of excluded matter (if it is, it falls within the 'as such' qualification and is excluded under Article 52(2) or Article 53); and,
4. check whether the contribution is technical in nature

The EPO in turn was highly critical of Lord Justice Jacob's four step test, arguing that it was 'not consistent with a good faith interpretation' of Articles 52 or 53.[91] Lord Justice Jacob's four part test was also questioned as appropriate for determining patent eligibility for computer implemented inventions in the *Symbian v Comptroller General of Patents*[92] series of cases. In the *Symbian Series*, the English High Court, and later the Court of Appeals, attempted to resolve the difference between UK IPO and EPO authority on software patents. The patent in question related to a method for accessing data using a dynamic link library (DLL) (which as discussed in Section 1.3.2 of Chapter One is an example of an object-oriented programming technique). DLLs provide a means of allowing programs running on the applicant's operating system to refer to a library of generic methods. Accordingly, DLLs provide a more efficient mechanism for operating hardware.[93]

Justice Patten of the High Court maintained that the four part test was the correct way of determining patent eligibility. However, Lord Justice Neuberger, giving the judgment of the Court of Appeal, argued that the third part of the UK IPO test (the contribution requirement) was substitutable for the EPO's technical contribution test.[94] In particular, his Honour held that an invention could be shown to demonstrate patentability where it included a method for

---

[90] *Aerotel Ltd v Telco Holdings Ltd Ors* [2006] EWCA Civ 1371, paragraph 40.

[91] *T-154/04* (15th November 2006) [2008] O.J. EPO ('DUNS LICENSING ASSOCIATES/Estimating sales activity') paragraphs 12 and 13.

[92] *Symbian Ltd v Comptroller General of Patents* [2008] EWHC 518 (Pat); *Symbian Ltd v Comptroller General of Patents* [2009] RPC 1.

[93] Sherif Yacoub, Hany Ammar and Ali Mili, 'Characterizing a Software Component' (Paper presented at *International Workshop on Component-Based Software Engineering* , 1999) 4.

[94] *Symbian Ltd v Comptroller General of Patents* [2009] RPC 1, paragraphs 46 and 48.

improving the efficiency of software running on a general purpose computer.[95] The subsequent cases of *HTC Europe Co Ltd v Apple Inc* and *Lantana v Comptroller General of Patents* appeared to have adopted Lord Justice Neuberger's conclusion. In other words, these cases treatedthe 'technical effect' and 'technical contribution' tests are virtually synonymous.[96] These decisions established that Articles 52 and 53 do not introduce a complete restriction on software patenting. Instead, the exclusions provide some scope for applicants to acquire software patents provided that they have some physical effect. This test appears to set patent eligibility in the UK with the standard established by the US Supreme Court in *Benson*.[97]

As with the UK, the German *Bundesgerichtshof* ('BGH') and *Bundespatentgericht* ('BPatG') jurisprudence has also clashed with the EPO regarding the correct test for determining patent eligibility for German patents. On the one hand, the BGH in *Sprachanalyseeinrichtung (Language analysis device)* initially applied the 'technical effect' test to determine the Articles 52 and 53 subject matter exclusions,[98]. On the other hand, in subsequent cases both the BGH and BPatG held that a software program that simply relied upon a general-purpose computer would not meet the necessary requirements for patent eligibility.[99] Instead, the BGH argued that the subject matter exclusions were to be determined by reference to the test for patent eligibility under German law. Specifically, this test involved considering whether the patent provides a technical solution to a technical problem based on technical considerations. In other words, an algorithm could be patentable if it involves a physical entity, such as a computer network or specific computer.[100] More recent decisions have, like Lord Justice Neuberger in *Symbian*, attempted to reconcile EPO authority with German patent authority. However, even patents which meet the technical solution test are routinely struck down for a lack of novelty or inventiveness.[101] Furthermore, the UK's

---

[95] *Symbian Ltd v Comptroller General of Patents* [2009] RPC 1, paragraph 5; Sylvia Song, 'Software patents - a Comparison of the UK/European and Australian Laws on Patentability' (2015) [102] *Intellectual Property Forum: journal of the Intellectual and Industrial Property Society of Australia and New Zealand* 52 53.

[96] *HTC Europe Co Ltd v Apple Inc* [2013] EWCA Civ 451; *Lantana v. Comptroller General of Patents, Design and Trade Marks* [2014] EWCA Civ 1463.

[97] Brendon Beheshti, 'Getting beyond Abstract Confusion: How the United Kingdom's Jurisprudence Can Aid in Developing an Analytic Framework for Patent-Eligibility in Light of Alice v. CLS Bank' (2014) 10(2) *Washington Journal of Law, Technology & Arts* 137–152 150.

[98] Bundesgerichtshof [German Federal Supreme Court], *Sprachanalyseeinrichtung (language analysing device) Case X ZB 15/98*,11th May 2002 reported in (2002) OJ EPO 8–9.

[99] Bundesgerichtshof [German Federal Supreme Court], *X ZB 22/07*,20th January 2009 reported in (2010) GRUR Int. 2009 528; Bundesgerichtshof [German Federal Supreme Court], *Dynamische Dokumentengenerierung*,22nd April 2010 reported in (2010) GRUR Int. 2010 1003.

[100] Bundesgerichtshof [German Federal Supreme Court], *X ZR 3/12 Routenplanung*,18th December 2012 reported in (2013) GRUR 2013 275.

[101] Jochen Bühling, 'The Protection of Computer-Implemented Inventions – An Analysis of the Latest Case Law of

decision to withdraw from the EU following a referendum in 2016 ('Brexit') may have a hitherto unsuspected impact on attempts to introduce uniform standards for patent eligibility.[102]

3.3.3 *Patentable Subject Matter in Australia and New Zealand*

With respect to the scope of patent subject matter eligibility, Australian case law has largely been inherited from UK case law, but has also been influenced by US case law. This congruence is in spite of the fact that there are key differences in the statutory tests used to assess the scope of patent eligibility. The High Court in *NRDC v Commissioner for Patents* ('*NRDC*'),[103] established the 'manner of manufacture' test inherited from English patent law was the appropriate test for determining patent eligibility. The High Court further held that what amounts to a manner of manufacture to be determined on a case by case basis. Subsequent courts have held that it encapsulates any subject matter that involves 'an artificially created state of affairs of utility in the field of economic endeavour'.[104] Applying the *NRDC* test to software patenting, IP Australia concluded that algorithms per se were not capable of being patented, but the application of algorithms to produce 'an improvement in the operation of or the use of a computer' was capable of being patented.[105]

There has been considerable inconsistency in more recent Australian case law as to whether this physical requirement is necessary.[106] In *IBM v Commissioner of Patents*,[107] Justice Burchett in the Federal Court held that the rendering of a curve on a computer screen had a useful physical effect. Further, Justice Burchett noted that the algorithm constituted a new and unconventional use of a computer, and therefore warranted patent protection. In *CCOM Pty Ltd v Jiejing Pty Ltd*,[108] the Full Federal Court considered an algorithm for converting English alphanumeric characters into Chinese characters. The Full Federal Court held that there was no additional requirement for software patents beyond the need for

---

the German Federal Court of Justice' (2013) 8(12) *Journal of Intellectual Property Law & Practice* 957–960 960.

[102] Thomas Jaeger, 'Reset and Go: The Unitary Patent System Post-Brexit' (2017) 48(3) *IIC - International Review of Intellectual Property and Competition Law* 254–285 269-280; Graeme B. Dinwoodie and Rochelle Cooper Dreyfuss, 'Brexit and IP: The Great Unraveling' (2017) 39() *Cardozo Law Review* 967–996 991-993.

[103] *National Research and Development Corporation v Commissioner of Patents* (1959) 102 CLR 252, 269.

[104] *CCOM Pty Ltd v Jiejing Pty Ltd* (1996) 21 FCR 260, 295.

[105] IP Australia, 'Manual of Practice and Procedure' (last update 1 August 2017), s. 2.9.2.7

[106] Benjamin J. McEniery, above n36; Jessica C. Lai, 'The Nebulous "Invention": From "Idea and Embodiment" to "Idea/Embodiment and Observable Physical Effects?"' in Antoinette Maget Dominicé and Jessica C. Lai (eds.), *Intellectual Property and Access to Im/material Goods* (Edward Elgar Publishing 2016)121 123-5.

[107] *IBM v Commissioner of Patents* [1991] FCA 625.

[108] *CCOM Pty Ltd v Jiejing Pty Ltd* (1996) 21 FCR 260.

creating 'an artificially created state of affairs of utility in the field of economic endeavour',[109] overruling Justice Burchett's test requiring the software to include an additional physical effect on a standard computer. Ann Monotti is highly critical of the Full Federal Court's decision in *CCOM*, noting that its interpretation of 'manner of manufacture' amounts to a guide rather than a strict test. Monotti further notes the High Court's judgement in *NRDC* stated that 'economic endeavour' was not designed as a means of endlessly expanding the scope of patentable subject matter.[110] As with the US tests for patentable subject matter, *CCOM* appeared to open the scope of software patent eligibility without providing any limits on the extent of patent eligibility.

These criticisms were reflected in the subsequent decision of the Full Federal Court in *Grant v Commissioner of Patents* ('*Grant*').[111] which concerned another patent on a business method. This patent had been filed as an innovation patent, Australia's second tier utility patent model described in Section 3.5.2 of this Chapter. Both the Commissioner of Patents and the Full Federal Court rejected the notion that a pure 'economic effect' was sufficient to establish patent eligibility. Instead, the Full Federal Court held that there needed to be some useful product flowing from the operation of a method for it to be a patent eligible invention.[112] This relaxation of the physicality requirement has been indicated McEniery's study of business method patents acquired from 2000 to 2009 found that from a sample of 200 patents filed during this period, 75 claimed either a mixture of physical and non-physical claims or entirely non-physical claims.[113] However, of the 49 patents in McEniery's sample filed after *Grant* (where the Full Federal Court confirmed that a physical element was required for patent eligibility), only three were granted and all three involved claims to physically-transformative computer software.[114]

More recently, in the *Research Affiliates LLC v Commissioner of Patents* ('*Research Affiliates*')[115] and *RPL Central v Commissioner of Patents* ('*RPL Central*')[116] series of cases have considered patent eligibility. Specifically, the Full Federal Court examined the patent eligibility of a computer implemented securities trading portfolio and computer implemented vocational enrolment system respectively. For both patents, the respective Federal Courts held

---

[109] *CCOM Pty Ltd v Jiejing Pty Ltd* (1996) 21 FCR 260, 283.

[110] Ann L. Monotti, 'The Scope of Manner of Manufacture under the Patents Act 1990 (CTH) after Grant v Commissioner of Patents Comment' (2006) 34(3) *Federal Law Review* 460–478 465-7.

[111] *Grant v Commissioner of Patents* (2006) 154 FCR 62.

[112] *Stephen John Grant* [2004] APO 11; *Grant v Commissioner of Patents* (2006) 154 FCR 62, paragraph 47; Ann L. Monotti, above n110, 474; Jessica C. Lai, above n106, 124.

[113] Benjamin J. McEniery, above n36, 19-20.

[114] Benjamin J. McEniery, above n36 19-20.

[115] *Research Affiliates LLC v Commissioner of Patents* [2014] FCAFC 150.

[116] *Commissioner of Patents v RPL Central Pty Ltd* [2015] FCAFC 177.

that where an alleged invention is implemented using a computer, that invention must have a useful physical result in relation to a material or tangible entity.[117] The Full Federal Court in *RPL Central* held that software and business method patents were not a 'new class' of patent claim that sat on the boundaries of patent eligibility (unlike gene patents). Accordingly, these patents did not warrant an investigation of the factors related to manner of manufacture. This analysis can be contrasted with that which occurred in *D'Arcy v Myriad*.[118] In *D'Arcy*, the High Court held that for new categories of invention, a number of factors would need to be considered to determine whether those inventions amounted to a manner of manufacture.[119] Chapter Four discusses these factors in the context of gene patents in greater detail. These factors are 'factors connected directly or indirectly to the purpose of the Act', and can include whether the invention could 'give rise to a large field of monopoly protection' with negative impacts on innovation or have a chilling effect on activities beyond the scope of the patent.[120] These standards are in addition to the requisite standard of 'manner of manufacture'.[121]

After holding they did not lie at the boundaries of existing judicial developments, the Full Federal Court in *RPL Central* addressed the patent eligibility of software patents more broadly. Specifically, the Full Federal Court held that there must be more than an abstract idea behind the patent. Further, the computer must be integral to the invention, rather than performing its usual function.[122] Finally, although not supplying precise guidelines, the claimed invention being technical in nature was held to be an indicator of patent eligibility.[123] Accordingly, *Research Affiliates* and *RPL Central* import into Australian patent law a test for software equivalent to the technical effect test described by the UK Court of Appeals in *Aerotel/Macrossan*.[124] In the aftermath of *Research Affiliates* and *RPL Central*, this new requirement is being enforced strictly and a number of software patent applications have been consequently denied.[125] However, as part of the recent appeal from *Encompass Corporation Pty Ltd v InfoTrack Pty Ltd*,[126] the Full Federal Court has unusually empanelled a five judge

---

[117] *Research Affiliates LLC v Commissioner of Patents* [2014] FCAFC 150, paragraph 114.

[118] *D'Arcy v Myriad Genetics Inc* (2015) 258 CLR 334, paragraph 28; *Commissioner of Patents v RPL Central Pty Ltd* [2015] FCAFC 177, paragraph 119.

[119] *D'Arcy v Myriad Genetics Inc* (2015) 258 CLR 334, paragraphs 5, 28.

[120] *D'Arcy v Myriad Genetics Inc* (2015) 258 CLR 334, paragraph 28.

[121] *Commissioner of Patents v RPL Central Pty Ltd* [2015] FCAFC 177, paragraph 115-9.

[122] *Commissioner of Patents v RPL Central Pty Ltd* [2015] FCAFC 177, 96.

[123] *Commissioner of Patents v RPL Central Pty Ltd* [2015] FCAFC 177, 98.

[124] James Scheibner and Dianne Nicol, 'Do Software Patents Inhibit Open Source Licensing in Australia?' (2015) 25(4) *Australian Intellectual Property Journal* 198.

[125] *Todd Martin* [2017] APO 33 (7 July 2017), paragraphs 43 to 49; *Discovery Life Limited* [2017] APO 36 (18 July 2017), paragraph 25.

[126] *Encompass Corporation Pty Ltd v InfoTrack Pty Ltd* [2018] FCA 421.

appeals bench. By contrast with the other courts that have ruled on software patentability, this court will have the potential to either confirm or overrule the previous decisions from *RPL Central* and *Research Affiliates*. Accordingly, it is expected that *Encompass Corporation Pty Ltd* will have a significant impact on patent eligibility.

In contrast to Australia, where case law has been used to shape the manner of manufacture requirement, amendments to the New Zealand *Patent Act 2013* (NZ) have brought New Zealand statutory standards in line with those found at the EPC. The Act now prevents patents on software 'as such' from being granted.[127] Overall, as Suzy Frankel and Jessica Lai note, the existence of the statutory restriction excluding patents as inventions in New Zealand imposes a restrictive threshold test. This threshold test is more difficult to satisfy than the equivalent threshold test in Australia.[128] However, in practice it is likely that the statutory exclusion on software patents in New Zealand will be subject to the same creative judicial interpretation as the Article 52 exclusion under EPC case law, as demonstrated by extensive EU case law.[129] Furthermore, there is evidence that patents that do meet the 'physicality' requirement mandated at both the judicial level in Australia and the statutory level in New Zealand with the as such requirement are being granted.[130] The fact that there has been such extensive litigation before the EPO Boards of Appeal, which over the past decade has reached a relative point of stability regarding the patent eligibility of computer programs, may provide the New Zealand IP office (along with New Zealand courts) with a roadmap to avoid the diversions that have plagued EU patent policy.[131]

Table 3.1 offers an overview of the current state of patent laws with respect to patentable subject matter across Australia, New Zealand, the EPO and the US.

3.4 COMPARATIVE LAW ON RESEARCH AND EXPERIMENTAL USE EXCEPTIONS

3.4.1 *The Narrow US Research Exceptions for Patent Infringement*

In contrast to the (historically) broad scope of patent eligibility in the US, a common law research exception for patent infringement has been construed narrowly. Such an exception would protect researchers conducting legitimate use with a patented product from a patent

---

[127] *Patents Act 2013* (NZ) s 11(3).

[128] Susy Frankel and Jessica C. Lai, *Patent Law and Policy* (LexisNexis NZ Limited, 2016) 362-3.

[129] Jessica C. Lai, above n106 124.

[130] AU 2013/263207B2, NZ702006A Method And System to Analyze Interference Susceptibility of a Radio Receiver Design

[131] Susy Frankel and Jessica C. Lai, above n128, 362.

Jurisdiction	Australia and New Zealand	Through the EPO	The US
Algorithms	Not patentable as a mathematical principle	Not patentable under Article 52(2)(c)	Not patentable as a mathematical principle
Software	*Australia* - Not patentable per se but can be patentable where it produces a useful, physical effect *New Zealand* - Not patentable per se but can be patentable when combined with a patent eligible invention.	Not patentable per se under Article 52(2)(c) but can be patented where there is a technical effect (EU) or where there is a technical contribution (UK).	Not patentable per se unless the remainder of the patent adds something beyond the unpatentable subject matter in software.
Computer Implemented Business Methods	Not patentable if merely implemented on a generic computer	Not patentable unless a technical effect is involved	Not patentable unless the abstract idea is transformed into a new and inventive patent eligible concept
Diagnostic Methods	*Australia* - possibly patentable if the factors from *D'Arcy v Myriad* call for the extension of patent law to a new class of patent claims	Patentable themselves under Article 53(3) (with patentability not extending to their uses)	Not patentable unless the invention has a specific 'therapeutic purpose' '

Table 3.1: Comparison of algorithm, software and business method patentable subject matter across Australia, New Zealand, the EU and the US.

infringement action. However, since the 1980s Federal Circuit decisions have confined the operation of this doctrine to a very narrow set of circumstances. These exceptions largely relate to resolving regulatory delay in the pharmaceutical context. This narrow scope is in spite of the theoretical legal and economic arguments in favour of expanding this research exception.[132] In particular, in the decision in *Duke University v Madey* ('*Madey*'),[133] the Federal Circuit interpreted any potential exception under US law narrowly. The Federal Circuit concluded that the use of a patented laser largely for research purposes by a university still had the financial benefit of attracting students to the faculty. In turn, the subsequent research outputs had the consequence of securing research funds, therefore falling outside the scope of the research exception. Katherine Strandburg argues the decision in *Madey* runs completely counter to the generally held perception in the scientific community that non-industrial or non-commercial university research does not constitute patent infringement.[134] This conflict may discourage researchers from engaging in research where either the potential for patent infringement or the cost of cross licensing of patented technologies is too great.[135]

In addition, there is longstanding uncertainty about the scope of reverse engineering of

[132] *Embrex, Inc. v Service Engineering Corp.* 216 F.3d 1343 (2000), 1352 per Judge Rader.

[133] *Madey v Duke University* 307 F3.d 1351 (2002), 1362.

[134] Katherine J. Strandburg, 'What Does the Public Get - Experimental Use and the Patent Bargain' (2004) [1] *Wisconsin Law Review* 81–156 83.

[135] Viola Prifti, above n6, 82.

software as discussed in Section 2.5 of Chapter Two.[136] Accordingly, the absence of an explicit research exception casts a long shadow over the reverse engineering of patented software.[137] This uncertainty may have a particularly singular negative effect on software development for research purposes in universities. This negative effect is due to the fact that software may have dual basic and applied research purposes. Katherine Strandburg has independently suggested a business method patent exception (equivalent to other technology specific patent exception such as an exception for plant breeders).[138] This patent exception would extend the operation of the prior user right defence under US patent law. The prior use defence initially encompassed prior use of business methods but was later amended to include a general prior use defence (via the *American Invents Act*).[139] This prior use defence immunises an alleged infringer who has not sought a patent but has nevertheless used the patented method prior to the patent being filed. However, a key weakness of these exceptions in the context of software would be the difficulty in identifying relevant software patents. It would also be necessary to determine whether the alleged infringing use would be protected by either a research exception or by prior use rights.[140] This vexing question is a difficulty that Chapters Five, Six and Seven will explore in greater detail from an empirical perspective. In particular, subsequent chapters will examine whether academic bioinformaticians perceive research exceptions and prior usage rights as a valuable strategy to protect their research.

3.4.2 *Experimental Use Exceptions in the EU*

Unlike the US, individual member states of the EU have developed a significantly broader range of exceptions for patent infringement. At the supranational level, there have been several attempts to introduce uniform rules on patent infringement. Article 27 of the *Agreement on a Unitary Patent Court (UPC)*[141], a draft agreement which has not been yet incorporated into the EPC, includes the following requirements:

> The rights conferred by a patent shall not extend to any of the following:
>
> (a) acts done privately and for non-commercial purposes;

---

[136] Brian Fitzgerald et al., 'Innovation, Software, and Reverse Engineering' (2002) 18(1) *Santa Clara High Technology Law Journal* 121 122.

[137] Katherine Strandburg, 'Patent Fair Use 2.0' (2011) 1(2) *UC Irvine Law Review* 265 279.

[138] Katherine J. Strandburg, 'What if There Were a Business Method Use Exemption to Patent Infringement' (2008) *Michigan State Law Review* 245–278 256.

[139] *35 U.S. Code §273 - Defense to infringement based on prior commercial use*; *American Inventors Protection Act 1999* (Public Law 106-113); *America Invents Act 2011* (Pub L No 112-129) section 5.

[140] Katherine J. Strandburg, above n138, 249.

[141] *Agreement on a Unified Patent Court* [2013] *OJ* C No 175 /56.

(b) acts done for experimental purposes relating to the subject-matter of the patented invention;

(c) the acts allowed pursuant to Article 13 of Directive 2001/82/EC[142] or Article 10 of Directive 2001/83/EC;[143]

...

(k) the acts and the use of the obtained information as allowed under Articles 5 and 6 of Directive 200924EC[144]

As mentioned previously, the future of this Agreement is somewhat in doubt due to the UK's decision to leave the EU. However, prior to this Agreement, nationally several EPC member states, including the UK, Germany, France and Belgium, have introduced broad research exceptions into their national patent legislation.[145] Despite the widespread existence of these research exceptions, there remains considerable deviation in how national courts and legislatures have treated the operation of these provisions.[146] For example, the Belgian Act exempts acts which are conducted with or on the patented invention if those acts are for 'scientific purposes'. This broad definition would appear to support both pure scientific and mixed commercial-scientific inquiries.[147] This interpretation is supported by a pair of German decisions, *Clinical Trials I* and *Clinical Trials II*. These decisions pertain to the equivalent exception in the *Patentgesetz.*[148] In *Clinical Trials I*, the German Federal Supreme Court held that with respect to list of activities protected under the exception:

> Since the provision makes no limit, either qualitative or quantitative, on the experimental acts, it cannot matter whether the experiments are used only to check the statements made in the patent or else to obtain further research results, and whether they are employed for wider purposes, such as commercial interests.[149]

---

[142] This article of this directive relates to pharmacological test data for veterinary medical products sold in the EU

[143] This article of this directive relates to pharmacological test data for human medical products sold in the EU

[144] This directive, as discussed in Chapter Two, pertains to the legal protection of computer programs in the EU, and defines the limits of copyright exceptions for error correction (Article 5) and developing interoperable products (Article 6)

[145] *Patents Act 1977* (UK) § 60(5), c. 37; *Code de la propriété intellectuelle 1992* (French Intellectual Property Code) art. L613-5; *Patentgesetz [German Patent Act] 1980* § 11; *Belgian Patent Act* art. 28(1)(b).

[146] Esther. van Zimmeren and Geertrui Van Overwalle, 'False Sense of Security Offered by Zero-Price Liability Rules? Research Exceptions in the US, Europe and Japan in an Open Innovation Context' in Ruth Okediji and Margo Bagly (eds.), *Global Perspectives on Patent Law* (Cambridge University Press 2011)379 405-6.

[147] Viola Prifti, above n6, 92.

[148] *Patentgesetz [German Patent Act] 1980* §11 No. 2.

[149] Bundesgerichtshof [German Federal Supreme Court], *Klinische Versuche (Clinical Trials) I*,30th October 1997 reported in (1997) RPC 623 639.

This perspective was confirmed by the Federal Supreme Court in *Clinical Trials II*:

> ...The purpose the experiment is intended to serve does not at all have to be of a purely scientific nature. According to this, the commercial orientation does not from the outset turn the experimental activity into an impermissible patent infringement. Something else will then have to determine when it is no longer a matter of the further elucidation of the conditions, effects, applicability, and producibility of the object of the invention, but of clarification of commercial facts such as the needs of the market, acceptance of prices, and possibilities of distribution.[150]

Applying this authority to reverse engineering and the development of interoperable software, it would appear these activities would be exempted by the experimental use exception if that software was being developed in a research context.[151] However, the UK position on the experimental use exception was narrowly interpreted in *Monsanto Co v Stauffer Chemical Co*.[152] Specifically, this exception only extended as far as the use of a patent to gather strictly experimental data; any data gathered for a commercial purpose was not protected. Moreover, there remain lingering doubts to the congruence of an experimental use exemption that extends to both uses with and on a patented invention under international law. In particular, the Belgian and German exceptions may fall foul of Article 30 of the TRIPS Agreement, which restricts any exceptions that interfere with the rights holders' rights. Both Chris Dent and Strandburg argue that extending the exception to research with a patented invention would adversely impact on the legitimate interests of those holding research tool patents. Accordingly, these exceptions would therefore be contrary to Article 30.[153] However, until these provisions are tested before the WTO Disputes Settlement Panel, it is unlikely that any definitive conclusion will be reached on the extent of these research exceptions.

3.4.3 *Research Exceptions in Australia and New Zealand*

In contrast to the US, and in much the same way as some EU jurisdictions, both Australia and New Zealand have introduced a statutory exemption on patent infringement for experimental

---

[150] Bundesgerichtshof [German Federal Supreme Court], *Klinische Versuche (Clinical Trials) II*,17th April 1998 reported in (1998) RPC 423 432.

[151] Ashwin van Rooijen, *The Software Interface Between Copyright and Competition Law: A Legal Analysis of Interoperability in Computer Programs* (Kluwer Law International, 2010) 102.

[152] *Monsanto Co v Stauffer Chemical Co* [1985] RPC 515.

[153] Katherine J. Strandburg, above n134, 121; Chris Dent, 'The TRIPS agreement and an experimental use exception for 'research tools" (2011) 44(1) *Australian Economic Review* 73–78 77.

purposes.[154] In New Zealand, the need for these exemptions grew from the common law exemptions that evolved in response to the needs of certain patent holders. These exceptions included regulatory review for pharmaceutical and agricultural researchers who needed regulatory approval from government agencies.[155] This reform resulted in the insertion of an experimental use exception into the new *Patent Act 2013* (NZ). Whilst Australian legislative development was also influenced by the presence of regulatory exceptions,[156] the political incentive for a broad research exception emerged from public controversies. These public controversies included the decision by Genetic Technologies Ltd seeking patent royalties from research institutes for use of genetic tests. These tests had been exclusively licensed from Myriad Genetics, a spin-off company from the University of Utah.[157] Simultaneously, the Australian Law Reform Commission (ALRC) released a report into the use of patents with respect to genetic technologies that recommended the introduction of an experimental use exemption for patents.[158] These factors culminated in the introduction of an experimental use exemption in the 2012 Australian amendments to the *Patent Act 1990* (Cth).

However, the boundaries of these experimental use exemptions remain untested in both Australia and New Zealand. Previous case law in New Zealand suggested that experimental use exemptions would not extend to the commercial use of technologies.[159] However, the new legislation does not contains this distinction. Instead, the new legislation focuses on whether a third party examining the claims could test the validity of the invention and the scope of the claims.[160] The abolition of this delineation between commercial and non-commercial use arguably broadens the potential of reverse engineering patented software using the source code. Further, this test could encompass applying that software in a way not previously envisaged by other researchers. Despite the benefits that the new experimental use exemption might bring, it suffers from the same limitations as a fair use defence for copyright with respect to digital research techniques. In other words, it is difficult to determine whether extending a software package or reusing it in another context would fall within the boundaries of an experimental use exemption. In the absence of case law to assist with this determination, subsequent chapters will assess how these determinations are made by researchers in an empirical context.

---

[154] *Patents Act 1990* (Cth) section 119C; *Patents Act 2013* (NZ) section 143.

[155] *Monsanto Co v Stauffer Chemical Co (No 1)* (1984) 1 NZIPR 518; *Smith Kline French Laboratories Ltd v Attorney-General (NZ)* (1991) 2 NZLR 560, 562, 566, 569.

[156] Ann L. Monotti, 'The Australian Experimental Use Exemption: A Current Overview' (2009) 12(5) *The Journal of World Intellectual Property* 422–445 423.

[157] Matthew Rimmer, 'The Freedom to Tinker: Patent Law and Experimental Use' (2005) 15(2) *Expert Opinion on Therapeutic Patents* 167–200 186.

[158] Australian Law Reform Commission, *Genes and ingenuity: gene patenting and human health*, No 99 (2004).

[159] *Monsanto Co v Stauffer Chemical Co (No 1)* (1984) 1 NZIPR 518, 533.

[160] Susy Frankel and Jessica C. Lai, above n128, 256-7.

Table 3.2 offers an overview of the current state of research exemptions across each of the three jurisdictions examined.

Jurisdiction	Australia and New Zealand	Through the EU	The US
Research Exemptions	Research permitted on the patented subject matter in both Australia and New Zealand but not with the patented subject matter	Research permitted on and with the patented subject matter in some jurisdictions (Belgium and Germany) but not others (such as the United Kingdom)	Research exemptions strictly limited to pure scientific inquiry only.

Table 3.2: Comparison of research exemptions in Australia and New Zealand, the EU and the US

3.5 UTILITY MODELS AND SUI GENERIS PROTECTION

3.5.1 *Utility Models in the EU and the Discordance in Second Tier Patent Systems*

A final interesting divergence between patent laws in the US and in EU member states (such as Austria, Denmark, Finland, France, Germany and Italy) is the presence of utility or second tier patents. These utility models are designed to protect incremental innovations as opposed to novel inventions in these European jurisdictions.[161] Although the TRIPS Agreement does not explicitly mention the existence of utility models, Article 2(1) does require that signatories give effect to the earlier Paris Convention. Article 1(2) of the Paris Convention provides that signatories may give effect to utility models.[162] In addition, Article 1 of the Strasbourg Agreement, which establishes an international patent classification system (which will be discussed in further detail in Chapters Five and Six), requires signatories to provide IPC classification for 'patents for invention... inventor's certificates, utility models, and utility certificates'.[163] These reforms were preceded by national reform in countries such as Germany which created utility models to resolve the overlap between copyright laws and designs with an industrial purpose.[164]

Whilst there is an international framework supporting the existence of utility models, it does not provide any minimum standards for such utility models. Further, efforts to harmonise

---

[161] Uma Suthersanen, *Utility Models and Innovation in Developing Countries* (International Centre for Trade and Sustainable Development (ICTSD), 2006) 1-2.

[162] *Paris Convention for the Protection of Industrial Property*, opened for signature 14th July 1967, 828 UNTS 305 (entered into force 26th April 1970) Article 1(2); *Marrakesh Agreement Establishing the World Trade Organization, annex IC, The Agreement on Trade Related Aspects of Intellectual Property Rights ('TRIPS Agreement')*, opened for signature 15th April 1994, 1867 UNTS 3 (entered into force 1st January 1995) Article 2(1).

[163] *Strasbourg Agreement Concerning the International Patent Classification*, opened for signature 21st March 1971, 1160 UNTS 483 (entered into force 7th October 1975) Article 1.

[164] Kelsey Martin Mott, 'The Concept of Small Patent in European Legal Systems and Equivalent Protection under United States Law' (1963) 49(2) *Virginia Law Review* 232–261 235-6.

utility model protection across Europe have largely failed, and so unlike the European patent system, the unitary patent system operates on the national level.[165] Despite significant deviation between the different national implementations of utility patent models, there are a number of common features of different utility patent models in Europe. These features usually include a reduced term of protection for utility model patents (on average between 7 to 10 years) and limitation of utility model protection to certain fields of technology. For example, utility models are limited to processes under the German and Austrian law and three dimensional objects under the Italian, Danish and Finnish laws.[166] The reduced scope of utility models is balanced by the lowered examination and administrative requirements, as well as lowered costs.[167]

The economic benefits of utility model patents are somewhat contentious. Given Germany's ranking as the sixth most innovative economy in the world, there has been significant interest in understanding the role that utility model patents play in this success.[168] In particular, the German utility model has been attributed with providing small to medium enterprises with a mechanism to protect incremental innovation, particularly in the sphere of electronic engineering.[169] The absence of a prior art searching requirement in the Belgium and Netherlands utility model systems meant that these rights were of limited or even negative economic value. This questionable economic value resulted in the abolition of these utility model systems in 2009 and 2008 respectively. This concern is reflected in Australia, where there are plans to abolish the utility model 'innovation patent system'.

3.5.2 *Innovation Patents in Australia*

Similar to the European jurisdictions described above, Australia also features an innovation patent system, which was introduced in 2000 to replace an existing petty patent system as part of major reforms to the existing *Patent Act*.[170] By contrast, no such utility model system exists within New Zealand at present. These innovation patents lack the substantive examination procedure associated with standard patents, but are available for a significantly shorter period of time of 8 years. As for the European utility models, the economic benefits of the Australian

---

[165] Hanns Ullrich, 'Expansionist Intellectual Property Protection and Reductionist Competition Rules: A TRIPS Perspective' (2004) 7(2) *Journal of International Economic Law* 401–430 411.

[166] *Gebrauchsmustergesetz [Utility Model Act] 2009* §1(2) and §2(3); Uma Suthersanen, above n161, 2.

[167] Timo Minssen, Berthold Rutz and Esther van Zimmeren, 'Synthetic Biology and Intellectual Property Rights: Six Recommendations' (2015) 10(2) *Biotechnology Journal* 236–241 239.

[168] Knut Blind, Jakob Edler and Michael Friedewald, *Software Patents: Economic Impacts and Policy Implications* (Edward Elgar Publishing, 2005) 70, 75; Uma Suthersanen, above n161, 16.

[169] Knut Blind, Jakob Edler and Michael Friedewald, above n168, 29; Andrea Bonaccorsi, Jane Calvert and Pierre-Benoit Joly, 'From Protecting Texts to Protecting Objects in Biotechnology and Software: A Tale of Changes of Ontological Assumptions in Intellectual Property Protection' (2011) 40(4) *Economy and Society* 611–639 619.

[170] *Patents Amendment (Innovation Patents) Act 2000* (Cth).

innovation patent system are highly contested. On the one hand, Andrew Christie and Sarah Moritz argue that the innovation patent system largely serves the function that it was designed for. In other words, utility models were designed to provide a system of protection for incremental innovation, particularly in fast moving industries such as software development.[171] On the other hand, a subsequent economic study by Benjamin Mitra-Kahn and others was unable to establish that the innovation patent system has any discernible macroeconomic impact on innovation. This lack of economic effect was due to underutilisation of the system by potential applicants.[172]

Table 3.4 offers an overview of the current status of utility models across each of the three jurisdictions considered.

Jurisdiction	Australia and New Zealand	Through the EPO	The US
Utility Models	Present as an 'innovation patent' but at risk of abolition, not present in New Zealand	Present in some EU jurisdictions, either as a short term patent equivalent or a patent limited to certain forms of technology	Not present in the US.

Table 3.3: Comparison of utility model protection across Australia, New Zealand, the EU and the US.

3.5.3 *A Sui Generis Regime For Software - An Alternative to the Patent System?*

As discussed in Section 2.2.2 of Chapter Two, there was initially significant resistance to the idea of protecting software through either copyright or patent law. In particular, the US strongly supported the idea of extending copyright protection to software. However, both the EU and Japan explored the possibility of introducing a sui generis right, or a novel regime to protect new innovation practices, for protecting software. The ostensible advantage of a *sui generis* regime is that it provides protection for emerging forms of technology where other, technology neutral forms of intellectual property protection do not provide adequate incentives for protection.[173] An example of a *sui generis* regime with supporting economic evidence for this argument is the plant breeders' right regime. This regime was developed in recognition of the fact that plants are self reproducing yet the cost of producing new plant varieties is relatively high. Accordingly, this right has been extensively used by plant breeders.[174] By way of comparison, an example of a *sui generis* regime which has not

[171] Sarah L. Moritz and Andrew F. Christie, 'Second-Tier Patent Systems: the Australian Experience' (2006) 28(4) *European Intellectual Property Review* 230 238.

[172] Benjamin Mitra-Kahn et al., 'The economic impact of innovation patents' (Economic Research Paper No 5, IP Australia, May 2015) 5.

[173] Megan Richardson, 'Sui Generis Intellectual Property Law Reform: Issues for Australia' (2001) 32(1) *Victoria University of Wellington Law Review* 19–46 20.

[174] Mercedes Campi and Alessandro Nuvolari, 'Intellectual Property Protection in Plant Varieties: A Worldwide Index (1961–2011)' (2015) 44(4) *Research Policy* 951–964 953.

provided adequate incentive for further research has been the semiconductor chipset protection regime. This regime was introduced into the TRIPS Agreement due to concerns about extensive patenting having a detrimental impact on semiconductor research and development.[175] However, as Thomas Hoeren notes, this *sui generis* regime has largely fallen into disuse and has, ironically, been replaced by extensive patent cross licensing.[176]

What justifications exist for protecting software using a *sui generis* regime? The primary concerns about the existing regimes extend to the uncertainty regarding the scope of functional protection for copyright. Proposals for sui generis regimes to replace software patents are driven by concerns regarding the cost of and ambiguity regarding the scope of patent rights for software. Most proposals for a sui generis regime for software have suggested decreasing the term of protection for software compared to that offered under copyright law from the life of the author plus seventy years a shorter term of ten to thirty years.[177] In addition, the majority of these *sui generis* proposals have also recommended incorporating the utility and limited examination requirements associated with second tier and utility model patents. At the same time, these proposals mandate a heightened requirement for establishing infringement.[178] Nevertheless, the greatest challenge for introducing any *sui generis* approach for software copyright will be to construct an economic case for its introduction. Although the EU and Japan eventually rejected adopting a *sui generis* mechanism for protecting software, at the national level *sui generis* protection was enacted in France for software and introduced into the *French Copyright Act* of 1985.[179] This regime was abolished with the introduction of the European Commission Software Directive, which standardised copyright protection for software across the European Union.[180] Further, there are concerns about whether such a regime would solve the negative externalities associated with intellectual property protection more effectively than existing systems.[181] Subsequent chapters of this thesis address each of

---

[175] *Marrakesh Agreement Establishing the World Trade Organization, annex IC, The Agreement on Trade Related Aspects of Intellectual Property Rights ('TRIPS Agreement')*, opened for signature 15th April 1994, 1867 UNTS 3 (entered into force 1st January 1995) Articles 35 to 38.

[176] Thomas Hoeren, 'The Semiconductor Chip Industry – The History, Present and Future of Its IP Law Framework' (2016) 47(7) *IIC - International Review of Intellectual Property and Competition Law* 763–796 792-3.

[177] Vikrant Narayan Vasudeva, 'A Relook at Sui Generis Software Protection Through the Prism of Multi—Licensing' (2013) 16(1-2) *The Journal of World Intellectual Property* 87–103 97-8.

[178] Laurence Diver, 'Would the Current Ambiguities Within the Legal Protection of Software be Solved by the Creation of a Sui Generis Property Right for Computer Programs?' (2008) 3(2) *Journal of Intellectual Property Law & Practice* 125–138 138.

[179] Jerome H. Reichman, 'Legal Hybrids between the Patent and Copyright Paradigms' (1994) 94(8) *Columbia Law Review* 2432 2481.

[180] *Council Directive 91/250 of 14 May 1991 on the Legal Protection of Computer Programs* [1991] *OJ* (1 122) No 42.

[181] Jane C. Ginsburg, 'Four Reasons and a Paradox: The Manifest Superiority of Copyright Over Sui Generis Protection of Computer Software' (1994) 94(8) *Columbia law review* 2559–2572 2560-2.

these issues in greater detail.

3.6 CONCLUSION

This chapter has charted the evolution of software patent protection in the US, the EU, Australia and New Zealand. Whilst the case law in this area has remained relatively undisturbed in the EU for at least a decade and a half, there has been significant upheaval in the US. This upheaval is particularly noticeable with respect to software and business processes patents. These changes limit the scope of patent eligibility for software related inventions. Therefore (in part), there has been a convergence between US precedent and European, Australia and New Zealand precedent on whether computer implemented methods are patentable. However, there is still some deviation between the tests that apply for determining how software inventions (as well as other intangible inventions) can be patented. In particular, the USPTO appears to be struggling with defining the two part Alice test so as to achieve the intended policy objective for removing overly broad patents.

As subsequent chapters discuss, reducing the scope of patentable subject matter for software may prevent the use of patents as a means of claiming an overly broad monopoly on public domain methods. However, by contrast, the scope of research exemptions in the US remain relatively strict relative to the EU, Australia and New Zealand, which permit research with and on a patented invention. The existence of a narrow research exemption for use of patented software has the potential to forestall open source development practices. As further chapters of this thesis will discuss, open source development is largely dependent on cumulative research and development involving software. Further, there have been attempts to create an alternative, *sui generis* form of protection for software. However, these forays have been stymied by concern about whether such regulatory changes would be necessary as opposed to superior private ordering strategies.

Chapter 4

APPLYING COMMON POOL RESOURCE AND KNOWLEDGE COMMONS THEORY TO OPEN SOURCE COMPUTATIONAL BIOLOGY

4.1 INTRODUCTION

Chapters Two and Three discussed the extent to which intellectual property laws (copyright and patent law respectively) can apply to software and data. This chapter now addresses the relationship between copyright, patent and *sui generis* regimes, and bioinformatics development (under both proprietary and open source regimes). Section 4.2 revisits the key analyses in Chapter Two and Three concerning how copyright and patent law apply to different elements of a bioinformatics workflow. These elements include bioinformatics software, annotated and unannotated sequence data and sequencing hardware, as discussed in Chapter One. Section 4.2 also considers whether copyright and patent protection have the potential to be either positive or negative factors in the success and failure of open source bioinformatics projects.

Continuing this analysis further, Section 4.3 addresses the existing literature on theoretical frameworks for understanding the development of open scientific data and research tools. It also addresses the theoretical frameworks developed for the analysis of open source software communities. Specifically, these frameworks include regulatory capitalism, social capital theory, and actor network theory. These frameworks also include Elinor Ostrom's Institutional Analysis and Design (IAD) and Katherine Strandburg, Michael Madison and Brett Frischmann's adapted Knowledge Commons framework. Section 4.3 finally explains why the Knowledge Commons framework represents the culmination of the previously mentioned theories. Accordingly, it justifies why the Knowledge Commons framework the most suitable framework for this study. Section 4.3 explains the utility of the Knowledge Commons framework for assessing the impact of copyright and patent laws on both the *reproducibility* and *sustainability* of open source bioinformatics projects.

Section 4.4 explores the role of commons-based governance for socio-technical resources. It discusses how the Knowledge Commons framework can be used to understand how socio-technical commons resources and governed. In particular, Section 4.4 considers what normative role intellectual property rights (namely patents and copyright) play in the governance of such socio-technical resources. Section 4.4 then examines the specific role that intellectual property rights can play in the governance of open source bioinformatics projects. These strategies range from open source and open data licensing based on copyright laws to patent pooling using the strategic disclosure of patents. This methodological comparison frames the empirical analysis in Chapters Five and Six.

4.2 POLICY ARGUMENTS ON PATENT AND COPYRIGHT PROTECTION FOR BIOINFORMATICS SOFTWARE

4.2.1 *Availability of Copyright Protection for Bioinformatics Software*

As alluded to in Chapter Two, the extent of copyright protection for bioinformatics largely depends on whether copyright is directed to the literary or non-functional parts of a software program. On the one hand, across all four jurisdictions considered in Chapter Two, the source code used for bioinformatics software is unquestionably protected by copyright law as a literary work. Accordingly, source code for sequence alignment algorithms and phylogenetic tree structures will be protected by copyright.[1] Although copyright offers limited protection for software functionality, in almost all jurisdictions it arises automatically without registration.[2] Therefore, copyright represents the most straightforward means for a developer to protect their work.[3] Within the field of bioinformatics, there are examples of public domain, open source and proprietary packages that all contain copyright licences. Some of the earliest bioinformatics algorithms, such as BLAST for sequence alignment and Fitch and Margoliash's phylogenetic tree algorithms, were released as public domain works without copyright restrictions.[4] By comparison, PAUP*, a software package that is used for drawing phylogenetic trees, is copyrighted and released under a proprietary licence for general use and a time limited trial licence.[5] Section 4.2.3 discusses these differences in licensing models in more detail.

On the other hand, there may be elements of bioinformatics software which are unlikely to fall within the realm of copyright protection in any of the jurisdictions under consideration.[6] First, a number of popular bioinformatics packages remain so because they provide graphical user interfaces. By way of comparison, both BLAST and PAUP* run via command line interfaces.[7] These command line interfaces may be flexible for academic computer scientists

---

1 Vincent J. Carey and Victoria Stodden, 'Reproducible Research Concepts and Tools for Cancer Bioinformatics' in Michael F. Ochs, John T. Casagrande and Ramana V. Davuluri (eds.), *Biomedical Informatics for Cancer Research* (Springer US 2010)149–175 170-72.

2 It should be noted that in the United States, copyright can be registered at the US Copyright Office.

3 Jorge L. Contreras and A. Jamie Cuticchia, *Bioinformatics Law: Legal Issues for Computational Biology in the Post-genome Era* (American Bar Association, 2013) 10.

4 Stephen F. Altschul et al., 'Basic local alignment search tool' (1990) 215(3) *Journal of Molecular Biology* 403–410; Joshua Elkington, *BLAST Algorithm. Contribute to elkoutwest/BLAST development by creating an account on GitHub* (April 2016) <https://github.com/elkoutwest/BLAST>.

5 James C. Wilgenbusch and David Swofford, 'Inferring Evolutionary Trees with PAUP*' (2003) 00(1) *Current Protocols in Bioinformatics* 6.4.1–6.4.28.

6 As discussed in Section 2.3 of Chapter Two with respect to the extent of software copyright to functional aspects of computer programs.

7 Where users enter in particular commands for the software to perform via a command line interface.

but not so intuitive for researchers who do not have experience with terminal software. In addition, command line tools may rely on particular parameters to operate, which can create complications for reproducibility.[8] However, packages such as Bioconductor provide graphical user interfaces for performing tasks such as identifying chromosomal locations.[9] As stated above, the underlying source code of Bioconductor may warrant copyright protection (as the literal element of the computer program). By contrast, the graphical user interface will only warrant copyright protection insofar as the creative expression of that interface is concerned.[10] In addition to graphical bioinformatics software, bioinformatics researchers have become increasingly reliant on application programming interfaces (APIs) to facilitate data sharing.[11] However, as discussed in Section 2.5 of Chapter Two, the decision of the Court of Appeals in *Oracle v Google* raises uncertainty as to whether APIs warrant copyright protection in the US. The uncertainty injected into the software copyright debate by *Oracle v Google* may have a consequent effect on open source licensing.

Further, there is significant contention as to whether copyright (and, by effect, open source licensing) can be applied to sequence data that has been annotated using bioinformatics software. As discussed in Section 2.4.4 of Chapter Two, copyright does not vest in mere compilations of data in the US, Australia and New Zealand. Further, Christopher Holman, Claes Gustafsson and Andrew Torrance note that the US Copyright Office has rejected extending copyright protection to engineered DNA sequences.[12] These restrictions on copyright raises the question whether the attribution and redistribution requirements under open source licences apply to data, as many open source licences are drafted for software.[13] In these jurisdictions, raw data collected using sequencing hardware would be treated as a compilation of facts, and would therefore likely fall outside the scope of copyright protection. Further, the *sui generis* right from the EU Database Directive became operational in 1996. At that time, most open source licences did not provide explicit protection for data.[14] There is

---

[8] Alexandre Hocquet and Frédéric Wieber, ‘’Only the Initiates Will Have the Secrets Revealed’: Computational Chemists and the Openness of Scientific Software’ (2017) 39(4) *IEEE Annals of the History of Computing* 40–58 52.

[9] Robert C. Gentleman et al., ‘Bioconductor: Open Software Development for Computational Biology and Bioinformatics’ (2004) 5(10) *Genome Biology* R80 R80.3.

[10] Omar Johnny, Marc Miller and Mark Webbink, ‘Copyright in Open Source Software - Understanding the Boundaries’ (2010) 2(1) *International Free and Open Source Software Law Review* 13–38 25-6.

[11] Pieter B. T. Neerincx and Jack A. M. Leunissen, ‘Evolution of Web Services in Bioinformatics’ (2005) 6(2) *Briefings in Bioinformatics* 178–188 179.

[12] Christopher M. Holman, Claes Gustafsson and Andrew W. Torrance, ‘Are Engineered Genetic Sequences Copyrightable?: The U.S. Copyright Office Addresses a Matter of First Impression’ (2016) 35(3) *Biotechnology Law Report* 103–111 111.

[13] *Feist Publications, Inc. v Rural Telephone Service Co.* 499 U.S 340 (1991); *IceTV Pty Ltd v Nine Network Australia Pty Ltd* (2009) 239 CLR 459.

[14] Simone Aliprandi, ‘Open licensing and databases’ (2012) 4(1) *International Free and Open Source Software Law*

also considerable uncertainty as to how the derivative work provisions apply to ensure that derivative databases are either relicensed or contain attributions to the original work.[15]

One past attempt to resolve this issue was the International HapMap Public Access Licence (the HapMap Licence). The HapMap Licence was designed as an equivalent to a restrictive open source licence for public access to human haplotype data gathered for the International HapMap project. Both academic and commercial operators could reuse HapMap Project data on the prerequisite that no patents were applied for using any data from the database.[16] Once contributors to the HapMap project had placed a sufficient amount of data in the public domain, the HapMap Licence itself was abandoned.[17] However, Donna Gitter notes that the key weakness of the HapMap Licence was its construction as contractual licence rather than a contractual copyright licence. As a result, only parties to the HapMap Licence were bound by the promise not apply proprietary rights to data that they had obtained from the Project.[18] Accordingly, third parties could acquire patents using HapMap Project data without abiding by the HapMap Licence if they could obtain that data from a HapMap licensee.[19]

Nevertheless, the use of open source licensing in bioinformatics is complicated by more recent technological advances. First, modern bioinformatics involves the use of a custom bioinformatics pipeline. A bioinformatics pipeline describes a process by which files are 'shepherded' through a series of transformations to process sequence data and add metadata to these files.[20] Before DNA or protein sequences can be analysed using bioinformatics software, the raw base pair sequences need to be identified.[21] This process usually occurs using either polymerase chain reaction (PCR) or next generation sequencing (NGS) platforms to map and convert the alignment data into a readable file format.[22] Each of these three points demonstrate the limitations of copyright protection as it extends to bioinformatics. That is, whilst software source code falls within the boundaries of copyright protection, and APIs may

---

Review 5–18 10-2.

[15] Simone Aliprandi, above n14, 12.

[16] International HapMap Consortium, 'International HapMap Project Public Access License 1.1' ([2003]) *PubRL* 4.

[17] The purpose of this database was to create a substantial field of prior art had been created to forestall any patent claims. (David Castle, *The Role of Intellectual Property Rights in Biotechnology Innovation* (Edward Elgar Publishing, 2009) 348)

[18] Donna M. Gitter, 'Resolving the Open Source Paradox in Biotechnology: A Proposal for a Revised open Source Policy for Publicly Funded Genomic Databases' (2006) 43(5) *Houston Law Review* 1475 1489.

[19] Arti Rai and James Boyle, 'Synthetic Biology: Caught between Property Rights, the Public Domain, and the Commons' (2007) 5(3) *PLoS Biology* 391-2.

[20] Jeremy Leipzig, 'A Review of Bioinformatic Pipeline Frameworks' (2017) 18(3) *Briefings in Bioinformatics* 530–536 530.

[21] As described in Section 1.4.1 of Chapter One.

[22] Craig Shimasaki, *Biotechnology Entrepreneurship: Starting, Managing, and Leading Biotech Companies* (Academic Press, 2014) 208.

be copyright eligible, graphical interfaces (where functional rather than literary) and compilations of data do not. Critically, as the next section discusses below, many NGS platforms and instruments are likely to be patented and protected by proprietary licensing.[23]

Secondly, artificial intelligence and machine learning algorithms play an increasingly important role in bioinformatics pipelines. As discussed in Chapters One and Two, simple artificial intelligence techniques like decision trees were already used for constructing phylogenetic trees. Machine learning expands on artificial intelligence algorithms by automating the processing of large, complex data sets. These tools have become increasingly important for predictive analysis in a wide variety of fields.[24] Machine learning algorithms can be particularly useful at analysing large sets of genomic, proteomic or metabolomic data that could not otherwise be analysed using standards statistical techniques. Amongst machine learning algorithms, unsupervised machine learning, such as clustering algorithms, operate without an existing set of labels, rather than imposing labels on data. By contrast, supervised machine learning learning algorithms rely on smaller training data sets, which then 'trains' the algorithm on how to process larger data sets. For example, a supervised machine learning algorithm that identifies the properties of particular genes using a training set of known genes.[25] However, because both supervised and unsupervised machine learning algorithms are shaped by the data they receive, there is an intermingling of data and software in these algorithms. In other words, the functionality and results produced by the algorithm will depend on the data that it receives. Accordingly, this combination of software and data will blur the boundaries between copyright in software and any potential copyright in data.[26] Some researchers have developed licences designed to cover both the software and the data used for machine learning, particularly where data comes from heterogenous sources.[27] However, the increased emphasis on functionality might drive academic and commercial bioinformaticians towards patent protection for artificial intelligence and machine learning algorithms in bioinformatics.

---

[23] Somak Roy et al., 'Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists' (2018) 20(1) *The Journal of Molecular Diagnostics* 4–27 5.

[24] David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn about Machine Learning' (2017) 51(2) *U.C. Davis Law Review* 653–718 655.

[25] Maxwell W. Libbrecht and William Stafford Noble, 'Machine learning applications in genetics and genomics' (2015) 16(6) *Nature Reviews Genetics* 321–332 322-3.

[26] Maarten Truyens and Patrick Van Eecke, 'Legal Aspects of Text Mining' (2014) 30(2) *Computer Law & Security Review* 153–170 169.

[27] Guido Governatori et al., 'One License to Compose Them All' in Harith Alani et al. (eds.), *The Semantic Web – ISWC 2013* (Springer Berlin Heidelberg 2013)151–166 152-3.

4.2.2 *Availability of Patent Protection for Bioinformatics*

Although there is limited case law on the patent eligibility of bioinformatics inventions, reference to analogous fields of subject matter may indicate the bounds of patentable subject matter. As revealed in Chapter Three, software which is tied to a specific hardware platform will likely satisfy the requirements for physicality, technical effect or manner of manufacture. This type of software could include that used to control Leroy Hood's automated sequencer. Accordingly, if the other requirements for patent eligibility are satisfied, these inventions amount to patentable subject matter in the four jurisdictions considered in Chapter Three. In this regard, an analogy can be drawn between bioinformatics patents and synthetic biology patents. The latter have been granted to academic and private sector patent applicants alike.[28] Whether synthetic biology patents will remain patent eligible subject matter largely depends on the jurisdiction in question.[29] In *Association for Molecular Pathology v Myriad Genetics*[30] the US Supreme Court invalidated Myriad's claims on isolated sequences. However, the Supreme Court upheld Myriad's patents on cDNA sequences because they were synthetically produced and not identical to the naturally occurring sequence.[31] By contrast, the Australian High Court rejected cDNA patent claims in *D'Arcy v Myriad*[32] on the grounds that the isolated DNA had a fundamentally informational character. Therefore, these patent claims were held not to amount to a manner of manufacture under Australian patent law.

The US decisions would prima facie appear to support the patent eligibility of synthetic biology patents, provided they extend beyond identifying biomarkers for naturally occurring DNA sequences. However, in Australia, *D'Arcy v Myriad* would appear to place isolated DNA sequences beyond the scope of patentable subject matter. Despite this outcome, the Australian Patent Office (APO) has upheld some synthetic biology patent claims in the aftermath of *D'Arcy v Myriad*, such as a patent on synthetic double stranded RNA in *Arrowhead Research Corporation*.[33] In this case, the crucial factor was that RNA is normally single stranded (as opposed to double stranded). Accordingly, the claimed invention was held to include more than just genetic information and therefore amounted to a manner of manufacture.[34] Outside of the US and Australia, Articles 52 and 53 of the European Patent Convention exclude animal

---

[28] Paul Oldham and Anthony Mark Cutter, 'Mapping Global Status and Trends in Patent Activity for Biological and Genetic Material' (2006) 2(2) *Genomics, Society and Policy* 62 83.

[29] Timo Minssen, 'Patenting Human Genes in Europe- And how it compares to the US and Australia' in Duncan Matthews and Herbert Zech (eds.), *Research Handbook on Intellectual Property and the Life Sciences* (Edward Elgar Publishing Limited 2017)26 29.

[30] *Association for Molecular Pathology v Myriad Genetics, Inc.* 133 S. Ct. 2107 (2013), 2119-20.

[31] *Association for Molecular Pathology v Myriad Genetics, Inc.* 133 S. Ct. 2107 (2013), 2119.

[32] *D'Arcy v Myriad Genetics Inc* (2015) 258 CLR 334, paragraph 6.

[33] *Arrowhead Research Corporation* [2016] APO 70.

[34] *Arrowhead Research Corporation* [2016] APO 70, paragraph 28.

and plant varieties from the scope of patentable subject matter.[35] However, the European Patent Office Board of Appeals in *Howard Florey Institute*[36] held that isolated DNA sequences would fall outside this exception as the isolation was the result of technical processes. Likewise, the UK Supreme Court in *Human Genome Sciences Inc v Eli Lilly Co*[37] held that an isolated DNA sequence was patentable as a technical process, provided that its industrial applicability was disclosed.[38] If a claimed synthetic biology invention involves isolating naturally occurring DNA or creating novel DNA sequences, it is likely to be considered patentable subject matter in the EU.[39]

The question of patentable subject matter becomes more tenuous when considering bioinformatics algorithms. In the US, the *Mayo* and *Alice* tests in unison would appear to exclude any patent on any software that simply implements a mathematical method for a general-purpose computer. This decision may exclude computer implemented methods or computer readable media claims, including biostatistics and biological database management software, from the scope of patent eligibility.[40] This conclusion is supported by patent validity actions before the USPTO Board of Appeals and Inferences (BPAI), which has denied bioinformatics patents.[41] These results suggest that whether a bioinformatics invention warrants patent protection in the US will largely be determined by whether that patent is directed to an abstract idea. If that patent is directed to an abstract idea, the question then turns to address whether there is something more that transforms that abstract idea into patent eligible subject matter. In the aftermath of *Alice*, Trent Ostler and Michael Gollin confirmed this finding by studying the bioinformatics patents granted by the USPTO. They noted these patents were characterised by either framing the claimed invention narrowly or tying the invention to a specific biological process. Osler and Gollin give the specific example of bioinformatics inventions for classifying particular types of hepatocellcular carcinoma tumours or quantitative assessment of cancer tumour growth for a particular patient.[42] Likewise, following the precedent from *Research Affiliates* and *RPL Central*, the Australian Patent Office (APO) has recently upheld a bioinformatics algorithm for testing diagnostic

---

[35] *Convention on the Grant of European Patents*, opened for signature 5th October 1973, 1065 UNTS 254 (entered into force 7th October 1977).

[36] *T0272/95* (23rd October 2002) [2002] 'Relaxin/Howard Florin Institute').

[37] *Human Genome Sciences Inc v Eli Lilly & Co* [2012] RPC 6.

[38] *Human Genome Sciences Inc v Eli Lilly & Co* [2012] RPC 6, 40-41.

[39] Timo Minssen, above n29, 33.

[40] Gregory J. Kirsch and Charlie F. Brown, 'Software patents' in Jorge L. Contreras and A. Jamie Cuticchia (eds.), *Bioinformatics Law: Legal Issues for Computational Biology in the Post-genome Era* (American Bar Association 2013) 54-5.

[41] *Ex Parte Kelker* (Tech Center 1600, No App. No. 2009-004635, (B.P.A.I., September 24 2010)).

[42] Trent Ostler and Michael Gollin, 'Which Types of Bioinformatics Inventions are Eligible for Patent Protection?' (2015) 21(2) *Journal of Commercial Biotechnology* 76–82 79-81.

devices in *Bio-Rad Laboratories*.[43] Before this decision, the APO upheld a patent on a bioinformatics algorithm for creating an animal feeding plan in *Mars v Hills Pet Nutrition*. The APO justified upholding this patent on the grounds this algorithm was analogous to a diagnostic method.[44] These cases would suggest that bioinformatics algorithms can be recognised as patentable subject matter in Australia.

In the EU, whether a bioinformatics algorithm will be patent eligible depends on whether the novel contribution lies in an excluded field under Articles 52(2) and (3) of the EPC. If the novel contribution lies in an excluded field, the patent will fall outside the realm of eligibility. This test was applied in *The Court of Edinburgh Napier University (Patent)*, where a digital forensic algorithm only involved the use of data mining techniques and was therefore excluded.[45] However, even where a bioinformatics patent is drafted as patent eligible subject matter, there is still potential for it to be invalidated due to failure to comply with the other requirements for patent eligibility. A patent was invalidated in *Beckman Coulter*,[46] a case before the EPO Board of Appeals. In this case, the Board of Appeals held that the patent claimed lacked sufficient description of the technical effect to constitute valid patentable subject matter. By contrast, in the subsequent case of *Perlin*,[47] the Board of Appeals held the novel contribution for the bioinformatics algorithm claimed was performed using dedicated laboratory equipment.[48] Accordingly, the patent was not directed towards excluded subject matter.

As for litigation, the recent *University of California v Broad Institute* dispute has loomed large over the computational biology landscape.[49] This case concerned priority of the CRISPR Cas-9 genome editing technology patents, and was resolved through settlement. Beyond this dispute, a search of US case law databases revealed only two other cases; *Illumina v Affymetrix*[50] and *DNA 2.0 v Genome Compiler Corporation*.[51] The first matter was summarily dismissed by a judge of the US District Court of Wisconsin. The second matter was settled through cross licensing of the algorithm in question. Outside the US, the only dispute involving a bioinformatics patent by academic developers, *Illumina and Others v Premaitha*

---

[43] *Bio-Rad Laboratories, Inc.* [2017] APO 24 (3 April 2018).

[44] *Mars, Incorporated v Hill's Pet Nutrition, Inc.* [2014] APO 67, 39-40.

[45] *The Court of Edinburgh Napier University (Patent)* [2011] UKIntelIP o16411, paragraphs 13-17.

[46] *T 0784/06* (23rd June 2010) [2010] 'Genotype determination/BECKMAN').

[47] *T 2050/07* (19th February 2013) [2013] 'DNA Mixture Analysis/Perlin').

[48] *T 2050/07* (19th February 2013) [2013] 'DNA Mixture Analysis/Perlin') 3.

[49] *Regents of the University of California v Broad Institute* (Fed. Cir., No No. 2017-1907, 2018).

[50] *Illumina, Inc. v Affymetrix, Inc.* (Dist. Court. WD Wisconsin, No Nos. 09-cv-277-bbc, 09-cv-665-bbc, 2010).

[51] *DNA Twopointo, Inc. v Genome Compiler Corporation* (ND. Cal., No No. 3:12-cv-06428-SI, 2013).

Health Plc,[52] involved a dispute between Stanford and the University of Hong Kong as co-claimants, and was quickly settled.[53] As Chapters Five and Six discuss later, this lack of litigation stands in contrast to the extensive litigation over the broader infringement of software patents and in particular university owned software patents.[54] The next section will address the complicating factors associated with open source licensing relative to proprietary licensing. It will also consider potential reasons as for why open sources licences or patents are not as heavily enforced in academic research.

4.2.3 *Proprietary Licensing and Open Source Licensing: Alternatives or Complements?*

No form of intellectual property is designed to provide perfectly optimal protection for a creative or inventive work.[55] However, the doctrinal analysis of copyright and patent protection[56] established that copyright and patent law overlap due to the dual literary-functional nature of software.[57] In addition, as alluded to in Section 2.5 of Chapter Two, the reverse engineering of software is tightly controlled through a combination of copyright, trade secrecy and contract law.[58] Through a veritable arms race of restrictive licensing and technological protection measures (TPMs), proprietary copyright licensing may weigh down the legitimate goals of copyright law and create major social costs. These social costs may be imposed not only on software users but also developers.[59] In part, the free and open source software community evolved in response to the software industry's increasing reliance on restrictive licensing.[60] By comparison, open source software licences themselves are relatively straightforward to interpret and are designed to be 'open textured'. This 'open textualism' allows for flexible interpretation in the face of new technology or new institutional

---

[52] *Illumina, Inc v Premaitha Health Plc* [2017] EWHC 2930 (Pat).

[53] Staff Reporter, *Preamitha Licenses NIPT Tech From Illumina Under Settlement* (19th September 2018) <https://www.genomeweb.com/molecular-diagnostics/premaitha-licenses-nipt-tech-illumina-under-settlement#.W-OM-3ozZcA>.

[54] Arti K. Rai, John R. Allison and Bhaven N. Sampat, 'University Software Ownership and Litigation: A First Examination' (2009) 87(5) *North Carolina Law Review* 1519–1570 1536-7.

[55] Shubha Ghosh, 'How to Build a Commons: Is Intellectual Property Constrictive, Facilitating, or Irrelevant?' in *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press 2007)209–245 222.

[56] In Section 2.3 of Chapter Two and Section 3.3 of Chapter Three

[57] Clark D. Asay, 'Copyright's Technological Interdependencies' (2015) 18(2) *Stanford Technology Law Review* 189–497 226.

[58] See (Ariel Katz, 'A Network Effects Perspective on Software Piracy' (2005) 55(2) *University of Toronto Law Journal* 155–216 156)

[59] Ariel Katz, 'Substitutions and Schumpeterian Effects over the Life Cycle of Copyrighted Works' (2008) 49(2) *Jurimetrics* 113–154 142-4.

[60] Christopher M. Kelty, *Two Bits: The Cultural Significance of Free Software* (Duke University Press, 2008) 192.

norms.[61] Nevertheless, there needs to be careful development of institutional policies to govern open source software use.[62] These specific requirements may change again within academic software development, as discussed in the next section.

4.2.4 *Proprietary Licensing and Open Source Licensing in Academic Research?*

Since the recognition of software patents, there has been ongoing debate as to how best balance the rights available for a patent holder.[63] As discussed in the introduction to this thesis, economist Michael Heller argues that collective action problems emerge where proprietary rights are too strong and proliferate amongst many entities. These rights collectively prevent individual actors from using a resource.[64] Heller has elaborated on the tragedy of the anti-commons emerging with respect to wireless broadcast spectrum ownership and airfield production in the US.[65] However, perhaps the most widely referenced application of this theory has been Heller and Rebecca Eisenberg's hypothesis regarding the potential for patents to stymie biomedical research.[66] The concerns over rent seeking effects or reach through licensing is more acute within academic research compared to privately funded research. This concern emerges from the fact that academic research has traditionally received public funding on the basis of research impact.[67] However, in the US, the *Bayh-Dole* Act radically changed university commercialisation practices by permitting US universities to acquire and hold patents for publicly funded research and exclusively licence the results in exchange for royalties.[68] Although the *Bayh-Dole* Act is arguably responsible for significant research commercialisation of publicly funded research, there are concerns that it may

---

[61] Chen Wei Zhu, "Copyleft' Reconsidered Why Software Licensing Jurisprudence Needs Insights from Relational Contract Theory' (2013) 22(3) *Social & Legal Studies* 289–308 292.

[62] Anna Kaarina Haapanen, 'Free and Open Source Software and the Mystery of Software Patent Licenses Under the GPL' (2015) 7(1) *International Free and Open Source Software Law Review* 19–28 21.

[63] Suzanne Scotchmer, 'Standing on the Shoulders of Giants: Cumulative Research and the Patent Law' (1991) 5(1) *The Journal of Economic Perspectives* 29–41 29-41; Thorsten Käseberg, *Intellectual Property, Antitrust and Cumulative Innovation in the EU and the US* (Bloomsbury Publishing, 2012) 10; Michael Noel and Mark Schankerman, 'Strategic Patenting and Software Innovation' (2013) 61(3) *The Journal of Industrial Economics* 481–520.

[64] Michael A. Heller, 'The Tragedy of the Anticommons: Property in the Transition from Marx to Markets' (1997) 111(3) *Harvard Law Review* 621–688 624.

[65] Michael Heller, 'The Tragedy of the Anticommons: A Concise Introduction and Lexicon' (2013) 76(1) *Modern Law Review* 6–25 8-9.

[66] Michael A. Heller and Rebecca S. Eisenberg, 'Can Patents Deter Innovation? The Anticommons in Biomedical Research' (1998) 280(5364) *Science* 698–701.

[67] Partha Dasgupta and Paul A. David, 'Toward a New Economics of Science' (1994) 23(5) *Research Policy* 487–521 496.

[68] *Bayh-Dole Act of 1980 (Patent and Trademarks Act Amendments of 1980)*; David C. Mowery and Bhaven N. Sampat, 'The Bayh-Dole Act of 1980 and University–Industry Technology Transfer: A Model for Other OECD Governments?' (2004) 30(1-2) *The Journal of Technology Transfer* 115–127 123.

incentivise reach through licences that place downstream restrictions on use and discourage swift scientific exchange.[69]

In contrast to the US, the EU does not have a unified legislative framework to drive technology transfer for universities and research institutes. In some EU countries patents vest with the individual inventors and universities are not involved in the commercialisation process. For example, in Germany, patents on publicly funded research are frequently assigned to state owned intermediate institutes.[70] In other EU jurisdictions there is greater industrial collaboration (such as the UK, which has been gradually moving towards a university ownership model).[71] Although neither Australia nor New Zealand have a uniform legislative framework for technology transfer, the *National Principles of Intellectual Property Management for Publicly Funded Research* are modelled on the *Bayh-Dole* Act. Accordingly, Australian universities behave in a largely uniform fashion with respect to technology transfer practices.[72] Concerns over the inflexibility of technology transfer offices have been reflected in how research communities structure themselves. Section 1.4.2 and 1.4.3 of Chapter One discussed how collaborative innovation influenced early progress in both computer science and molecular biology research and continues to influence their progress. For example, the academic software sharing culture drove the development of early networking and operating systems software. Likewise, the Human Genome Project and International HapMap project was largely achieved through the co-operation of government funded research agencies. Much of this collaboration was driven by private sector competition and concerns about patent acquisition on express sequence tags (ESTs).[73] As a possible consequence, it appears that Heller and Eisenberg's predictions of a widespread anti-commons in biomedical research have not eventuated. In particular, a series of studies suggest patents present fewer impediments to biomedical research than anticipated.[74]

---

[69] Anthony D. So et al., 'Is Bayh-Dole Good for Developing Countries? Lessons from the US Experience' (2008) 6(10) *PLOS Biol* e262 2078-80.

[70] Peter Tinnemann et al., 'Patenting of University and Non-University Public Research Organisations in Germany: Evidence from Patent Applications for Medical Research Results' (2010) 5(11) *PLOS ONE* e14059 6.

[71] Joaquın M. Azagra-Caro, Luis Plaza-Gómez and Ana Romero-de-Pablos, 'The origin of public research organisation patents: an economic approach' (2007) 16(4) *Research Evaluation* 271–282 272; Aldo Geuna and Federica Rossi, 'Changes to University IPR Regulations in Europe and the Impact on Academic Patenting' (2011) 40(8) *Research Policy* 1068–1076 1070.

[72] Michelle Marie Lockhart, Zaheer-Ud-Din Babar and Sanjay Garg, 'Drug development and research in New Zealand: policies affecting the industry' (2012) 73(1) *Drug Development Research* 1–10 6; Eleanor Flening, *30 Years After the Bayh-Dole Act: Rethinking the Australian Research Commercialisation Experience* (PhD Thesis, 2012) 138 <https://digitalcollections.anu.edu.au/handle/1885/9055>.

[73] Kenneth G. Huang and Fiona E. Murray, 'Entrepreneurial Experiments in Science Policy: Analyzing the Human Genome Project' (2010) 39(5) *Research Policy* 567–582 572-4.

[74] Rebecca S. Eisenberg, 'Noncompliance, Nonenforcement, Nonproblem - Rethinking the Anticommons in Biomedical Research Patent Law in Perspective' (2008) 45(4) *Houston Law Review* 1059–1100 1061.

To this end, Dianne Nicol and Jane Nielsen suggest that the question of the anti-commons does not rest with *what* intellectual property rights are available for biotechnology research, but more *how* those rights are used and *whom* they are used against.[75] Nicol and Nielsen suggest that any reform with respect to university ownership of patents or patent grants should be carefully directed to avoid devaluing the patent grant or freedom of contract for the cross licensing of patents. For example, few firms would be willing to invest in biotechnology research without patent protection as an incentive. By comparison, the rationalisations to patent publicly funded software are not as strong.[76] These lessened incentives are largely due to the ease with which software can be transported and reproduced, as well as the shorter lifespan of software. In addition, there is some empirical evidence to suggest that the institutional measures required to balance patenting and publication are not as mature for computer science as they are for biotechnology research. Victoria Stodden and Isabel Reich's study of patent-publication pairs for software related inventions at US universities suggested researchers do not publish software algorithms they later patent. Accordingly, these competing interests raise the question of how to balance this compromise as part of policy.[77] In addition, within computational science research there are examples of patenting negatively impacting research collaboration. One case involves the software package Gaussian, which had previously been released as an open source tool but was then re-released under a restrictive licence.[78] The authors of Gaussian released it under an ambiguous proprietary licence. This licence seemingly prohibited the inclusion of Gaussian output in publications. Accordingly, here were significant concerns about the reproducibility of results produced using Gaussian, and amongst computational chemists use of Gaussian became a significant bar on collaboration.[79] The next section discusses theoretical frameworks that can be used to examine the compromise between scientific software as a public good and as a private product.

---

[75] Dianne Nicol and Jane Nielsen, 'Australian Medical Biotechnology: Navigating a Complex Patent Landscape' (2005) 27(9) *European Intellectual Property Review* 313–318 318.

[76] Arti K. Rai, John R. Allison and Bhaven N. Sampat, above n54, 1523.

[77] Isabel Rose Reich and Victoria C. Stodden, 'Software Patents as a Barrier to Scientific Transparency: An Unexpected Consequence of Bayh-Dole' (Paper presented at *The Seventh Annual Conference on Empirical Legal Studies (CELS 2012)* , November 2012) 17-8.

[78] Jim Giles, 'Software company bans competitive users' (2004) 429(6989) *Nature* 231–231.

[79] Alexandre Hocquet and Frédéric Wieber, above n8, 50-1.

4.3 THEORETICAL FRAMEWORKS FOR ANALYSING OPEN SOURCE AND OPEN SCIENCE COMMUNITIES

4.3.1 *Competing Frameworks for Understanding Knowledge Sharing in Bioinformatics*

This section now examines the different theoretical frameworks that have been used to understand open source software development. In particular, this section briefly describes four of the main competing theoretical frameworks for understanding collaborative governance. Although somewhat competing, each of these frameworks build on top of and borrow from one another. These frameworks include Regulatory Capitalism, Social Capital Theory, Network Theory and the Knowledge Commons framework. Each of the key points of each theory are described in Table 4.1.

	Regulatory Capitalism	**Social Capital Theory**	**Network Theory**	**Knowledge Commons**
Key Concepts	Regulation occurs outside legal rules through administrative measures	Social relationships explain the collaboration and exchange between parties	Different relationships influence individual behaviour between parties	Institutional rules govern collective action of shared sources
Actors and elements	Regulatory state and members of regulated capitalist society	The relationships between individuals constitute a valuable resource for the conduct of social affairs	Different types of relationships influence individual behaviour	Rules adapted to the attributes of the community and the resource type under consideration

Table 4.1: A brief overview of the four competing theoretical frameworks for understanding collaborative research in scientific research.

Firstly, regulatory capitalism is concerned with the use of administrative measures via rule making and monitoring to enforce social governance. David Levi-Faur, one of the key regulatory capitalism scholars, uses telecommunications and electricity networks as examples of markets that rely on combined free market and regulatory mechanisms to succeed.[80] Within this framework, Levi-Faur notes that both patent and copyright laws have significant utility as regulatory mechanisms. Janet Hope, writing in concert with Dianne Nicol and John Braithwaite, further describes how regulatory capitalism can be used to influence intellectual property reform. In particular, Hope, Nicol and Braithwaite demonstrate how a combination of regulatory strategies and 'open' business models can be used to implement 'open source' biotechnology regimes in Australia.[81] However, as Hope notes, regulatory capitalism is fundamentally targeted at encouraging more inclusive approaches in public-private research

[80] David Levi-Faur, 'Regulatory Capitalism: The Dynamics of Change beyond Telecoms and Electricity' (2006) 19(3) *Governance* 497–525 503-4.

[81] John Braithwaite, Dianne Nicol and Janet Hope, 'Regulatory Captialism, Business Models and the Knowledge

collaborations.[82] Accordingly, as John Asuda and Christopher Yansell note, regulatory capitalism is effective at modelling market dynamics in infrastructure networks, or trade in or between nations.[83] By contrast, as Section 1.4.3 of Chapter One discussed, private-public research collaborations do not encapsulate all forms of bioinformatics research. Further, regulatory capitalism might not encapsulate the individual transactions that occur between participants in a research community. Therefore, a more definitive methodological framework is required to understand open source bioinformatics communities.

Secondly, network theory builds examines the specific types of connections that participants in a particular community have with one another. Specifically, network theory postulates that the position of actors within a network influences how they interact with other actors, and how that network forms.[84] For example, network theory has been used to examine our different human and non-human actors interact with each other in the establishment of cyberinfrastructure.[85] Further, examining commercial bioinformatics research firms that acquire patents, Bruce Rasmussen uses network theory to explain how these firms access complementary technologies.[86] Outside of longitudinal firm based interactions, network theory can also be used to assess how actors interacted with one another during a particular event. For example, Sebastian Haunss uses network theory to illustrate how proponents and opponents of patenting interacted with one another during the European debate over software patents.[87] However, network theory does not necessarily encapsulate all of the interactions that occur in an open source system.

Thirdly, social capital theory treats the relationship between participants in a community as a valuable resource for conducting social affairs.[88] In other words, social capital represents the benefits that accrue from individuals interacting with one another. Within academic

---

Economy' in John Braithwaite (ed.), *Regulatory Capitalism: How it Works, Ideas for Making it Work Better* (Edward Elgar 2008) 128-9.

[82] Janet Hope, 'An Introduction to Open Source Biotechnology' in Christopher Arup and William van Caenegem (eds.), *Intellectual Property Policy Reform: Fostering Innovation and Development* (Edward Elgar 2009) 148-9.

[83] John Kojiro Yasuda and Christopher Ansell, 'Regulatory capitalism and its discontents: Bilateral interdependence and the adaptability of regulatory styles' (2015) 9(2) *Regulation & Governance* 178–192 188.

[84] Walter W. Powell et al., 'Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences' (2005) 110(4) *American Journal of Sociology* 1132–1205 1140.

[85] David P. Randall, E. Ilana Diamant and Charlotte P. Lee, 'Creating Sustainable Cyberinfrastructures' (Paper presented at *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* , 2015) 1761.

[86] Bruce Rasmussen, *Creating and capturing value in the biopharmaceutical sector* (PhD Thesis, Victoria University, 2008) 205-6 <http://www.vu.edu.au/research>.

[87] Sebastian Haunss, *Conflicts in the Knowledge Society: The Contentious Politics of Intellectual Property* (Cambridge University Press, 2013) 122-3.

[88] Paul S. Adler and Seok-Woo Kwon, 'Social Capital: Prospects for a New Concept' (2002) 27(1) *The Academy of Management Review* 17–40 19.

scientific research, social capital can be best represented by citation rates. In other words, researchers who receive more citations for published academic work will receive greater recognition within their field, as well as funding from public and private sources. However, in a scientific context, social capital can also include trust and networks to guarantee research for mutual benefit.[89] For example, Jeffrey Furman and Scott Stern use social capital to explain data sharing by scientific researchers in biotechnology research as a strategy to encourage reciprocal citation. This reciprocal citation amplifies the impact of individual citations.[90] Likewise, in open source software communities, social capital is embodied by the recognition and reciprocity that flows from contribution to particular projects.[91] This concept of recognition constituting social capital links to the case law discussed in Chapter Two, where US courts recognised the non-monetary value of recognition under open source licences. Carolin Haeussler, in an empirical study of social capital, notes that information sharing is characterised as socially beneficial. However, her own study into social capital could not identify whether data sharing by itself had long term benefits to those sharing the data.[92] Further, a study by Sebastian Engelhardt and Andreas Freytag examine various effects for supply side contributions to open source. Their findings suggest that open source projects with strong intellectual property protection feature increased supply side contributions.[93]

Haeussler further argues that these frameworks should also address the question of the sustainability of a socio-technical commons resource. In this context, sustainability can be conceptualised in one of two ways; resource sustainability or functional sustainability. Resource sustainability refers to the ability of a particular resource to keep producing a particular good.[94] Sustainability can also be conceptualised as functional sustainability, or the ability of the governors of a commons to maintain its integrityThis integrity might can challenged by both external and internal pressures, as described in Section 1.5 of Chapter

---

[89] Benjamin Six et al., 'Trust and social capital in the design and evolution of institutions for collective action' (2015) 9(1) *International Journal of the Commons* 154.

[90] Jeffrey L. Furman and Scott Stern, 'Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research' (2011) 101(5) *American Economic Review* 1933–1963 1960-1.

[91] Rebeca Méndez-Durón and Clara E. Garcıa, 'Returns from social capital in open source software networks' (2009) 19(2) *Journal of Evolutionary Economics* 277–295 292.

[92] Carolin Haeussler, 'Information-Sharing in Academia and the Industry: A Comparative Study' (2011) 40(1) *Research Policy* 105–122 114.

[93] Sebastian v. Engelhardt and Andreas Freytag, 'Institutions, culture, and open source' (2013) 95() *Journal of Economic Behavior & Organization* 90–110 105.

[94] Steven A. Moore, *Pragmatic Sustainability: Theoretical and Practical Tools* (Routledge, 2010) 20-23; Daniel H. Cole, 'Learning from Lin: Lessons and Cautions from the Natural Commons for the Knowledge Commons' in Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press 2014)45–68 48; Tom Dedeurwaerdere, *Sustainability Science for Strong Sustainability* (Edward Elgar Publishing, 2014) 32-34.

One.[95] In addition, the three previous frameworks do not particularly focus on how researchers operate as a community.[96] It is for these reasons that this thesis relies on the Knowledge Commons framework. Crucially, as discussed in Section 4.4, the Knowledge Commons framework has been extensively used to study both scientific data sharing arrangements and open source software communities. Accordingly, there is a large body of literature which can be compared with the findings from this study. Before addressing the Knowledge Commons framework, the next section describes its progenitor, the Institutional Analysis and Design (IAD) Framework, in greater detail.

4.3.2 *Analysing Institutional Rules and Commons Pool Resources*

The Knowledge Commons theory found its genesis in the work of Elinor Ostrom into commons governance.[97] Commons governance is typically characterised by the lack of a central authority holding property rights and governing resource management.[98] Ostrom's early investigatory work revealed that many inland fisheries were not governed through private property ownership or state control, but instead by the fishermen themselves. These fishermen had created their own rules for sustaining their fisheries.[99] Guided by the emerging field of game theory, Ostrom developed a set of seven *design principles* to help govern socio-ecological resources.[100] These design principles are described in further detail in Table 4.2.

Ostrom's objective in establishing this rule classification system was to develop a broader framework for comparing the governance of different socioecological common pool resources, known as the IAD Framework.[101] Once the pertinent rules are classified, the IAD

---

[95] Steven A. Moore, above n94 20-23; Nicholas Matthew Weber, *A Framework for Analyzing the Sustainability of Peer Produced Science Commons* (PhD Thesis, University of Illinois at Urbana-Champaign, 2015) 29 <https://www.ideals.illinois.edu/handle/2142/88062>; John Anderies and Marco Janssen, *Sustaining the commons* (Arizona State University, 2016).

[96] Ann Johnson, 'Modeling Molecules: Computational Nanotechnology as a Knowledge Community' (2009) 17(2) *Perspectives on Science* 144–173 146.

[97] Amy R. Poteete, Marco Janssen and Elinor Ostrom, *Working Together: Collective Action, the Commons, and Multiple Methods in Practice* (Princeton University Press, 2010) 32-33.

[98] Yochai Benkler, 'Between Spanish Huertas and the Open Road: A Tale of Two Commons?' in Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press 2014)69–98 69.

[99] Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, 1990) 145.

[100] Elinor Ostrom, above n99, 91; Larry L Kiser and Elinor Ostrom, 'The Three Worlds of Action: A Metatheoretical Synthesis of Institutional Approaches' in Michael Dean McGinnis (ed.), *Polycentric Games and Institutions: Readings from the Workshop in Political Theory and Policy Analysis* (University of Michigan Press 2000) 184-205.

[101] Amy R. Poteete, Marco Janssen and Elinor Ostrom, above n97 32-33; Daniel H. Cole, above n94, 42.

1. Clearly defined boundaries.	There must be rules which clearly delineate the boundaries of the resource itself as well as what rights individuals have to withdraw resources from the total resource
2. Proportional equivalence between benefits and costs.	Members of the group must negotiate a system that rewards members for their contributions.
3. Collective-choice arrangements.	Group members must be able to create at least some of their own rules and make their own decisions by consensus. This design principle is supported by the assumption that individuals will contribute harder towards goals that they have agreed upon.
4. Monitoring.	Managing a commons is inherently vulnerable to free-riding and active exploitation that may undermine the long term sustainability of the resource.
5. Graduated sanctions.	Appropriators who violate operational rules are likely to be assessed graduated sanctions. These sanctions do not necessarily need to be serious at first and can include informal sanctions, but will escalate in severity where there is no compliance.
6. Conflict-resolution mechanisms	Appropriators have rapid access to low cost arenas to resolve conflicts.
7. Minimal recognition of rights to organise.	The rights of appropriators to devise their own institutions are not challenged by external government authorities.

Table 4.2: Elinor Ostrom's seven design principles for analysing common pool resources. Described in Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, 1990), 145 and David Sloan Wilson. Elinor Ostrom and Michael E Cox, 'Generalizing the core design principles for the efficacy of groups' (2013) 90(Supplemental) *Journal of Economic Behaviour & Organization S21, S22*

framework can be then used to test the effectiveness of these rules in governing commons based resources.[102] The IAD framework is divided into three levels of analysis:[103]

1. Exogenous variables, which are the inputs into a commons governance system, and include physical attributes of the shared resource(s), attributes of the community that manages that resource, and structural rules.

2. An Action Arena space where exogenous variables and actors interact, cooperate and conflicts emerge. This level of the IAD studies the stakeholders and resources of the commons as inventoried by the exogenous variables in a particular period of interaction. Using the understanding gained from cataloging the exogenous variables, the analyst observes this process and then identifies tangible outcomes, objective results, and the way that conflicts from the past or the present are voiced, contested, settled or deferred.

3. The outcomes and evaluation flowing from the analysis of exogenous variables in the action arena. Essentially, this process involves assessing how well an institution did or did not solve a collective action problem.

[102] Charlotte Hess and Elinor Ostrom, 'A Framework for Analysing the Microbiological Commons' (2006) 58(188) *International Social Science Journal* 335–349 347.

[103] Michael J. Madison, Brett M. Frischmann and Katherine J. Strandburg, 'Constructing Commons in the Cultural Environment' (2009) 95(4) *Cornell Law Review* 657–710 678-680; Brett M. Frischmann, 'Two enduring lessons from Elinor Ostrom' (2013) 9(4) *Journal of Institutional Economics* 387–406 8.

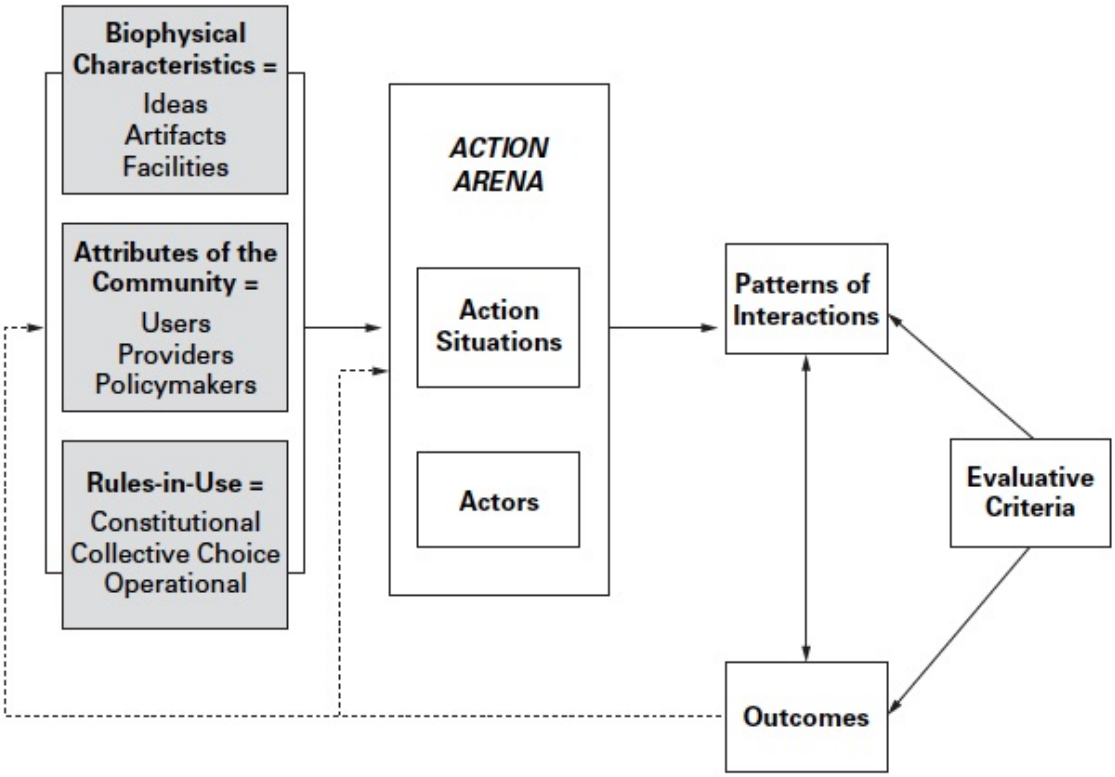

Figure 4.1: Biophysical, community and institutional characteristics within the IAD (From Elinor Ostrom and Charlotte Hess, 'A Framework for Analysing the Knowledge Commons' in Charlotte Hess and Elinor Ostrom (eds) *Understanding Knowledge as a Commons: From Theory to Practice* (The MIT Press, 2007), 41, 46. Reproduced with permission.)

The relationships between these different levels of analysis are demonstrated in Figure 4.1.

The rules that can be used to govern a commons based resource include both formal, legally enforced rules ('rules in form') and informal, community policed rules ('rules in use').[104] Using this categorisation approach, it is possible to compare the effectiveness of particular rules in promoting the sustainability of a common pool resource.[105] In the socio-ecological systems literature, these rules in use have been reinterpreted into boundary, position, informational, authority, aggregation, scope and payoff rules, which each correlate with Ostrom's original seven design principles.[106] Before turning to address considerations for the sustainability of socio-technical commons based resources, it is first necessary to address some of the important measures of success in the IAD framework.

---

[104] Elinor Ostrom, Roy Gardner and James Walker, *Rules, Games, and Common-pool Resources* (University of Michigan Press, 1994) 41-42; Elinor Ostrom, 'Background on the Institutional Analysis and Development Framework' (2011) 39(1) *Policy Studies Journal* 7–27 17.

[105] Michael D. McGinnis, 'An Introduction to IAD and the Language of the Ostrom Workshop: A Simple Guide to a Complex Framework' (2011) 39(1) *Policy Studies Journal* 169–183 170.

[106] Michael D. McGinnis, above n105, 174.

4.3.3 *Polycentricity, Sustainability and the Commons*

As the IAD framework became more widely adopted, Ostrom and other commons scholars began to recognise the heterogenous, complex nature of different commons. These commons had governance arrangements that extended beyond homogenous governance structures that might exist in a community.[107] The purpose of these governance structures was to guarantee sustainability, or the long term viability of the commons. In limited circumstances, Ostrom acknowledged that a central 'large-scale supportive institution' could perform the governance functions required to sustain the monitoring of a commons. In particular, Ostrom gave the example of the US Geological Survey, which supports disseminating information on geological surveys.[108] However Ostrom expressed scepticism regarding the effectiveness of large-scale government funded organisations for monitoring and enforcing commons property rights.[109]

Accordingly, polycentricity, or the adoption of many centres of decision making, may be preferable as a means of governance to ensure sustainability.[110] Within the IAD framework, polycentricity may operate on a number of different scales. Firstly, polycentric governance mechanisms may be multi-level in that they combine local, national and global units of governance. Secondly, polycentric governance mechanisms may be multi-type insofar that they include general purpose nested jurisdictions (such as traditional federalism) and specialised jurisdictional units. Finally, polycentric governance may be multi-sectoral, involving public, private, voluntary, community-based or hybrid organisations.[111] As Ostrom and Schweik both independently note, the ultimate goal of polycentricity is to maintain the internal sustainability and integrity of the commons.[112]

The concept of polycentricity demonstrates that property rights are not the only consideration in designing a governance framework. In particular, John Anderies argues that purely focusing on property rights as a mechanism for managing a commons can lead to misleading results in explaining how a commons is governed.[113] For example, property rights are less valuable when dealing with mobile resources such as fish as opposed to static

---

[107] Elinor Ostrom, 'Beyond Markets and States: Polycentric Governance of Complex Economic Systems' (2010) 100(3) *The American Economic Review* 641–672 644.

[108] Elinor Ostrom, *Understanding Institutional Diversity* (Princeton University Press, 2005) 278.

[109] Elinor Ostrom, above n99, 10.

[110] Elinor Ostrom, 'Policies that Crowd out Reciprocity and Collective Action' in Herbert Gintis (ed.), *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (MIT Press 1st Editioned, 2005)253 270; Elinor Ostrom, above n107, 643; Charles M. Schweik and Robert C. English, *Internet Success: A Study of Open-Source Software Commons* (MIT Press, 2012) 40,72.

[111] Michael D. McGinnis, above n105, 171.

[112] Elinor Ostrom, above n110 270; Charles M. Schweik and Robert C. English, above n110 40,72.

[113] John Anderies, Marco Janssen and Edella Schlager, 'Institutions and the performance of coupled infrastructure systems' (2016) 10(2) *International Journal of the Commons* 496.

resources such as forests.[114] As both later sections of this chapter and this thesis explore, outside of socio-ecological systems open socio-technical systems can also be characterised by their polycentric governance because of the different resources accompanied by each project, as well as the different roles that actors take.

4.3.4 *The Differences between Socio-Ecological and Socio-Technical Systems*

The key difference between socio-ecological and socio-technical systems depends on the types of resources that are managed as part of the commons. For environmental resources managed in a socio-ecological system, the following rules generally hold true:

1. resource boundaries are clear;
2. resource systems are usually stable, self contained and easy to observe;
3. solving social dilemmas is of high importance to appropriators (who might include funders, land users, governments and voluntary groups);
4. because of the sustainability of resources, the institutions and sanctions that govern those resources have developed over a long period of time and are equally stable.[115]

The resources that form part of socio-technical systems are marked by instability and unclear boundaries. This instability can be seen in the development of open source software.[116] In particular, the rise of network infrastructure has enabled developers to easily share source code with one another.[117] This networking effect means that open source software can rapidly evolve, in many cases much more radically than software developed in a traditional proprietary environment, due to the large group of loosely connected developers who contribute to a project.[118]

This networking effect in turn raises two further distinctions between socio-ecological and socio-technical systems.[119] Firstly, socio-technical systems are purposefully engineered rather

---

[114] Jacopo Baggio et al., 'Explaining Success and Failure in The Commons: The Configural Nature of Ostrom's Institutional Design Principles' (2016) 10(2) *International Journal of the Commons* 418.

[115] Charlotte Hess and Elinor Ostrom, 'A Framework for Analyzing the Knowledge Commons' in Charlotte Hess and Elinor Ostrom (eds.), *Understanding Knowledge as a Commons: from Theory to Practice* (MIT Press 2007) 45-46.

[116] As Section 1.4.2 of Chapter One illustrated.

[117] Charles M. Schweik and Robert C. English, above n110, 3-4.

[118] Jesus M. Gonzalez-Barahona et al., 'Macro-level Software Evolution: A Case Study of a Large Software Compilation' (2009) 14(3) *Empirical Software Engineering* 262–285.

[119] Nicholas Matthew Weber, above n95, 18-19 <https://www.ideals.illinois.edu/handle/2142/88062>.

than naturally occurring. The IAD framework requires a consideration of all social and ecological factors to understand how a socio-ecological commons is best sustained. Accordingly, the analysis of a socio-technical system must also take into account the purpose which that socio-technical system serves.[120] Secondly, the purposefully engineered nature of a socio-technical system blurs the lines between traditional user, producer and consumer roles.[121] One of the social dilemmas that may arise from the blurred distinctions between users, producers and consumers involves obfuscating the chain of ownership or authorship in a socio-technical system. This blurring effect can in turn have consequences for determining who holds copyright and patent protection.[122] This problem returns to the research questions at the heart of this thesis. The next section addresses how the knowledge commons framework can be used to model these interactions.

4.3.5 *Madison, Frischmann and Strandburg's 'Knowledge Commons' Framework*

The Knowledge Commons framework, is an adaptation of the IAD framework to analyse commons based resources composed of 'cultural' or socio-technical artefacts.The Knowledge Commons framework was initially referred to by Michael Madison, Brett Frischmann and Katherine Strandburg as the 'constructed cultural commons' framework as a means to understand cultural resources.[123] The purpose of the Knowledge Commons framework is to systematise investigations of socio-technical phenomena, develop a rigorous set of evaluative criteria and compare the findings of cases examining sharing arrangements in different socio-technical systems.[124] The Knowledge Commons framework shares the scalability of the IAD framework depending on the complexity of the socio-technical system under examination. However, the Knowledge Commons framework differs in a number of key ways from the IAD framework.

First, the Knowledge Commons framework recognises the interrelationship between attributes, resources and structural governance rules. The interrelationship between each of these exogenous variables represents the engineered nature of the resource governed in a

---

[120] Jeremy Pitt and Ada Diaconescu, 'Structure and Governance of Communities for the Digital Society' (Paper presented at *2015 IEEE International Conference on Autonomic Computing* , July 2015) 280-1.

[121] Yochai Benkler, 'Peer Production and Cooperation' in Johannes M. Bauer and Michael Latzer (eds.), *Handbook on the Economics of the Internet* (2016) 99.

[122] Jerome H. H. Reichman and Paul F. Uhlir, 'A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment' (2003) 66(1/2) *Law and Contemporary Problems* 315–462 323.

[123] Michael J. Madison, Brett M. Frischmann and Katherine J. Strandburg, above n103, 659.

[124] Brett Frischmann, Michael Madison and Katherine Strandburg, 'Governing Knowledge Commons – Introduction & Chapter 1' in Brett M. Frischman, Michael Madison and Katherine Strandburg (eds.), *Governing Knowledge Commons* (2014) 2,44.

socio-technical knowledge commons. These variables also represent the participation of users in the creation of this resource. For example, open source software development may be undertaken by both dedicated developers as well as standard users who wish to modify the software for their own ends.[125] Secondly, the user-innovation effect for open socio-technical systems is represented by collapsing the outcomes event described in the IAD framework into the patterns of interaction.[126] The merger of these two levels of analysis is representative of the fact that in the user innovation process, users can amend the socio-technical system to suit their needs. Finally, there may be complex and interconnected relationships between participants and resources within the Knowledge Commons framework. Accordingly, there is a blurring of the boundaries between the users and producers of resources associated with the commons.[127] The next section will address the role that intellectual property rights (specifically patents and copyright) play in the Knowledge Commons framework.

4.4 THE ROLE OF INTELLECTUAL PROPERTY RIGHTS IN THE KNOWLEDGE COMMONS FRAMEWORK

4.4.1 *Intellectual Property Theory and Knowledge Commons Governance*

Sections 4.2 illustrated the boundaries of intellectual property rights for bioinformatics. Section 4.3 then justified the choice of the Knowledge Commons framework for this thesis. This section examines the role that intellectual property rights (namely copyright and patent protection) play in commons-based governance. This analysis demonstrates *how* formal and informal norms emerge in bioinformatics communities around the normative constraints of patents and copyright laws.

The arguments surrounding the role that patents and copyright play in commons management for socio-technical resources are largely reflective of the broader debate about balancing the scope of patent and copyright protection.[128] However, caution must be exercised in attempting to use commons theory to impose normative judgements onto how patent and copyright law should 'best' operate.[129] A persistent criticism of intellectual property reform has been the tendency of legislators to expand existing patent and copyright scope, and introduce new intellectual property laws. This expansion then shrinks the scope of the public

[125] Charles M. Schweik and Robert C. English, above n110, 35.

[126] Michael J. Madison, Brett M. Frischmann and Katherine J. Strandburg, above n103, 682; Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press, 2014).

[127] Nicholas Matthew Weber, above n95, 57 <https://www.ideals.illinois.edu/handle/2142/88062>.

[128] Brett M. Frischmann, 'An Economic Theory of Infrastructure and Commons Management' (2004) 89(4) *Minnesota Law Review* 917–1030 933-4; Shubha Ghosh, above n55, 209-10.

[129] Daniel H. Cole, above n94, 46.

domain, which has been treated as roughly analogous to commons based initiatives.[130] However, focusing purely on the potentially restrictive aspects on a knowledge commons of intellectual property overstates the legalistic impact of intellectual property laws whilst ignoring its behavioural and institutional dimensions of a knowledge commons.[131]

The alternative perspective on intellectual property laws argues that patent and copyright laws are necessary to incentivise research and development, whilst ultimately benefiting the broader community through a limited monopoly.[132] For patent law, this disclosure occurs through the written description and enablement requirements, along with the patented invention becoming a public domain work following the expiration of the patent.[133] Although there are no formal registration and disclosure requirements for copyright, it encourages authors and artists to publish their works. Publication then makes these works more accessible to the wider community.[134] In the context of software engineering, both copyright and patent law may (imperfectly) serve these functions in providing some protection against reverse engineering of software source code.[135] Nevertheless, Ghosh argues that an overemphasis on the facilitative effects of intellectual property leads to an equivalent overemphasis on the importance of the role of free market economics and the notion that exclusivity offered by patents and copyright drives the free market. This overemphasis on free market economics in turn downplays the role of non market-based initiatives in the development of new technology.[136]

For example, Brian Love's comprehensive survey of attitudes towards patenting amongst professors at US computer science faculties demonstrates that some academic computer scientists regard patents as an alternative to publications. However, Love's study also suggested that interviewed computer scientists were highly disengaged from the patent acquisition and technology transfer process. More than half reported that they did not know whether or what percentage of royalties faculty staff received from patents.[137] Ghosh concludes that intellectual property rights should be seen as a means for constructing a

---

[130] James Boyle, *The Public Domain: Enclosing the Commons of the Mind* (Yale University Press, 2008) 160-1.

[131] Shubha Ghosh, above n55, 216.

[132] Thorsten Käseberg, above n63, 10-11.

[133] Shubha Ghosh, 'Patents and the Regulatory State: Rethinking the Patent Bargain Metaphor after Eldred' (2004) 19(4) *Berkeley Technology Law Journal* 1315 1330-1339.

[134] Ruth Towse, 'The Quest for Evidence on the Economic Effects of Copyright Law' (2013) 37(5) *Cambridge Journal of Economics* 1187–1202 1193-4.

[135] Pamela Samuelson and Suzanne Scotchmer, 'The Law and Economics of Reverse Engineering' (2001) 111() *Yale Law Journal* 1575–1664 1607-26.

[136] Shubha Ghosh, above n55, 219.

[137] Brian J. Love, 'Do University Patents Pay Off-Evidence from a Survey of University Inventors in Computer Science and Electrical Engineering' (2013) 16(2) *Yale Journal of Law and Technology* 285 314-8.

socio-technical commons rather than an end unto themselves.[138] Whilst this proposed model may encapsulate the perspectives of traditional academic researchers, it may come into conflict with the perspective of academic institutions. In addition, it may sit poorly with researchers who see the potential to produce industrially useful research.[139] The next section of this thesis will turn to address the different studies that have relied on the Knowledge Commons framework in studying socio-technical systems.

4.4.2 *The 'Bermuda Principles' and Genomic Data Sharing*

One area of genomics that has been subject to significant attention from legal and institutional economics scholars is the data sharing principles that formed during and in the aftermath of the Human Genome Project (HGP). When Human Genome Services, Celera and other private firms began to acquire patents, the leadership of the National Human Genome Research Institute (NHGRI) began to fear the prospect of patents impeding human genomics research.[140] In February 1996, leaders of the major genome sequencing centres met in Bermuda to discuss what principles should govern the sharing of genomic sharing.[141] The eventual agreement, known as the 'Bermuda Principles', attempted to forestall excessive patent claims on the human genome by mandating the rapid release (that is, a 24 hour release cycle) of human genome sequence data into public databases. The Bermuda Principles were designed to ensure that patents were only available on applied inventions that used, rather than directly claimed, particular human genome sequences.[142] This rapid release was made possible by the fact that, as discussed in Section 1.4.2 of Chapter One, all of the initial major sequencing centres had time sharing systems connected to GenBank.[143]

The Bermuda Principles were hailed by much of the scientific community for formalising norms governing the open exchange of proteomic and genomic data.[144] However, the emergent technology transfer offices (TTOs) at major research institutes, particularly in the US, were concerned that the principles contradicted the *Bayh-Dole Act*. The NHGRI

---

[138] Shubha Ghosh, above n55 219.

[139] Carolin Haeussler and Jeannette A. Colyvas, 'Breaking the Ivory Tower: Academic Entrepreneurship in the Life Sciences in UK and Germany' (2011) 40(1) *Research Policy* 41–54 43.

[140] As discussed in Section 1.4.3 of Chapter One

[141] Hallam Stevens, 'The Politics of Sequence: Data Sharing and the Open Source Software Movement' (2015) 50(4) *Information & Culture: A Journal of History* 465–503 465-466.

[142] Hallam Stevens, *Life Out of Sequence: A Data-Driven History of Bioinformatics* (University of Chicago Press, 2013) 242.

[143] Jorge L. Contreras, 'Bermuda's Legacy: Policy, Patents, and the Design of the Genome Commons' (2011) 12(1) *Minnesota Journal of Law, Science & Technology* 61 82.

[144] Jalayne J. Arias, Genevieve Pham-Kanter and Eric G. Campbell, 'The Growth and Gaps of Genetic Data Sharing Policies in the United States' (2015) 2(1) *Journal of Law and the Biosciences* 56–68 60.

responded to the potential for conflict by including an exception within the Bermuda Principles allowing participating research institutes to seek patents that have 'significant evidence of utility'. However, the NHGRI also noted that it would actively monitor 'grantee activity in this area' to determine 'whether or not attempts are being made to patent large blocks of primary human genomic DNA sequence'.[145]

To encourage researchers to submit data without fear of it being 'scooped' and used by other researchers in publications,[146] the Bermuda Principles were modified in 2000 to prohibit the use of data for 'large scale publications'.[147] The introduction of the Bermuda Principles also had a significant impact on the development of data sharing arrangements for other genomics research projects.[148] These projects include the International HapMap Project, the International Cancer Genome Consortium, the Malaria Genomic Epidemiology Network (MalariaGEN) and the International Series Adverse Events Consortium.[149] Victoria Stodden, Peixuan Guo and Zhaokun Ma note that these early data sharing initiatives in bioinformtics drove the relatively high rate of data and supplementary materials release in computational biology journals compared to other computer science journals.[150]

The impact of the Bermuda Principles is also relevant from an institutional economics perspective. In particular, the NHGRI's intellectual property policy and the Bermuda Principles articulate two of Ostrom's existing design principles; monitoring and graduated sanctions.[151] Furthermore, as a final draft sequence of the human genome was completed, the role of the NHGRI in managing data deposits also changed.[152] Jorge Contreras notes that the NHGRI initially had a funder/generator role insofar that it was responsible for generating and

---

[145] National Human Genome Research Institute, *Intellectual Property of Human Genomic Sequences* (9 May 2012) National Human Genome Research Institute <https://www.genome.gov/10000926/intellectual-property-of-human-genomic-sequence/>

[146] Peter Lee, 'Centralization, Fragmentation, and Replication in the Genomic Data Commons' in Katherine J. Strandburg, Brett M. Frischmann and Michael J. Madison (eds.), *Governing Medical Knowledge Commons* (Cambridge University Press 2017) 50.

[147] National Human Genome Research Institute, *NHGRI Policy for Release and Database Deposition of Sequence Data* (21 December 2000) National Human Genome Research Institute <https://www.genome.gov/10000910/policy-on-release-of-human-genomic-sequence-data-2000/>

[148] Rachel A. Ankeny and Sabina Leonelli, 'Repertoires: A post-Kuhnian perspective on scientific change and collaborative research' (2016) 60(Supplement C) *Studies in History and Philosophy of Science Part A* 18–28 24.

[149] Jorge L. Contreras, Aris Floratos and Arthur L. Holden, 'The International Serious Adverse Events Consortium's Data Sharing Model' (2013) 31(1) *Nature Biotechnology* 17–19 17-8.

[150] Victoria Stodden, Peixuan Guo and Zhaokun Ma, 'Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals' (2013) 8(6) *PLOS ONE* e67111 7.

[151] Referring back to Ostrom's seven design principles for a common pool resource in Section 4.3.2.

[152] Jorge L. Contreras, 'Leviathan in the Commons' in Katherine J. Strandburg, Brett M. Frischmann and Michael J. Madison (eds.), *Governing Medical Knowledge Commons* (Cambridge University Press 2017) 27-28.

managing data. Over time its role changed towards a regulatory role in the management of genomic data. The NHGRI now oversees the curation of data added to GenBank, cleaning data uploaded by researchers to ensure that it is of sufficient quality to be reused in downstream research.[153] In addition, the NHGRI is responsible for ensuring that this data collection process is compliant with national privacy and research ethics laws.[154] In the US, these laws include the *Health Insurance Portability and Accountability Act* of 1996 and the *Health and Human Service Regulations* 45 CFR part 46 (the 'Common Rule'). The NHGRI also ostensibly continues to manage the enforcement of the intellectual property policy governing genomic data deposited in GenBank.[155] Rather than formal action for copyright or patent infringement, the NHGRI (through its parent organisation, the NIH) instead attempts to resolve noncompliance through mediation, developing a remedial plan for compliance or, as a last resort, threatening funding cuts against the non-compliant centre.[156]

However, Jorge Contreras observes that out of 25,000 data access requests to GenBank between 2007 and 2015, only 27 of these access requests amounted to a policy violation in terms of data reuse.[157] As Contreras notes, this relatively low reporting rate could indicate that the policy is being infringed more frequently than anticipated by the NHGRI without being reported.[158] In the alternative, Peter Lee suggests that the lack of an explicit enforcement mechanism originates from the fact that the NHGRI explicitly structured GenBank to avoid conflicting intellectual property rights inhibiting public access to the genomic data stored within.[159] In part, this lack of enforcement capacity may be attributable to the difficulty in seeking formal intellectual property rights[160] once contributors renounce registered patent rights.

It is for this reason that other genomics databases have instituted policies that are less

---

[153] Sabina Leonelli, 'Valuing data in postgenomic biology : how data donation and curation practices challenge the scientific publication system' in Sarah S. Richardson and Hallam Stevens (eds.), *Postgenomics : perspectives on biology after the genome* (Duke University Press 2015)126–149; Nadine Levin and Sabina Leonelli, 'How Does One "Open" Science? Questions of Value in Biological Research' (2016) *Science, Technology, & Human Values* 0162243916672071.

[154] Robert Cook-Deegan, Rachel A. Ankeny and Kathryn Maxson Jones, 'Sharing Data to Build a Medical Information Commons: From Bermuda to the Global Alliance' (2017) 18(1) *Annual Review of Genomics and Human Genetics* 389 6.10, 6.17.

[155] Jorge L. Contreras, above n152, 30-31.

[156] Jorge L. Contreras, above n152, 31.

[157] National Institute of Health, *Genomic Data Sharing (GDS) - Categories, Statistics and Summary Information on Policy Violations (1 July 2015) Office of Science Policy <https://osp.od.nih.gov/scientific-sharing/categories-statistics-and-summary-information-on-policy-violations/>*

[158] Jorge L. Contreras, above n152, 32.

[159] Peter Lee, above n146 50.

[160] as well as copyright, due to the absence of copyright protection for data

oriented towards a governance regime of 'selective access and exclusivity'.[161] Geetrui van Overwalle gives the example of the Genetic Association Information Network (GAIN). GAIN included a contractual requirement preventing publication or downstream reuse of submitted data after a specific embargo period. In addition, GAIN required researchers to register with the GAIN Data Access Committee (DAC).[162] Of course, as the empirical research in this thesis will explore, the question of genomic data sharing arrangements is only part of the puzzle in the development of shared biological knowledge resources. In particular, sequence annotation and curation driven software can play a somewhat underemphasised role in the generation of new genomic knowledge.[163]

4.4.3 *Open Source Software Development as a Socio-Technical Commons*

Prior to Frischmann, Madison and Strandburg, Schweik observed that open source software development shares many of the characteristics of socio-ecological commons management.[164] However, rather than focusing solely on the role that intellectual property rights (namely patents and copyright) play in the development of open source software, Schweik and his collaborators focused on what factors underpin the success and failure (or abandonment) of open source software projects. To this end, Schweik and others seek to understand how open source developers solve the collective action problem of collaboratively developing software.[165] Crucially, Schweik and English note that open source software projects may vary significantly in size. These projects can include those with a large user base (such as the Linux kernel) to projects with a relatively small user base such as dedicated scientific softwareranging from projects with millions of lines of code and a large user base (such as the Linux kernel) to projects with a relatively small user base such as dedicated scientific software..[166] Schweik and English identify the key measures of success for open source projects as the number of releases for the project and the number of downloads. The number of releases, or revisions of the source code, represents whether the software is actively being developed. By contrast, the number of downloads represents whether the project is actively being used by its intended user base. In addition, Schweik and English distinguish between the

---

[161] Peter Lee, above n146, 46.

[162] Teri A. Manolio et al., 'New Models of Collaboration in Genome-Wide Association Studies: the Genetic Association Information Network' (2007) 39(9) *Nature Genetics* 1045–1051 1048-9.

[163] Peter Lee, above n146, 56-66.

[164] Charles M. Schweik, 'An Institutional Analysis Approach to Studying Libre Software Commons' (2005) 6(3) *Upgrade: The European Journal for the Informatics Professional* 17–27 21.

[165] Charles M. Schweik et al., 'Brooks' Versus Linus' Law: An Empirical Test of Open Source Projects' (Paper presented at *Proceedings of the 2008 International Conference on Digital Government Research* , 2008) 424; Charles M. Schweik and Meelis Kitsing, 'Applying Elinor Ostrom's Rule Classification Framework to the Analysis of Open Source Software Commons' (2010) 2(1) *Transnational Corporations Review* 13–26.

[166] Charles M. Schweik and Robert C. English, above n110, 57.

different stages of an open source project life cycle, depending on whether the project is in its initiation phase or in a growth phase.[167]

During the growth phase for a project, financial support for development is associated with the success of an open source project.[168] Tied to this are the motivating factors for producing source code. As later chapters of this thesis will explore, whether this motivation holds true for all open source projects, particularly in academic research, remains uncertain. In particular, a key factor motivating the production of scientific software is contributing to research practice and academic literature. The most commonly used measure to assess this impact is through citation.[169] In addition to citation, technology transfer and licensing of research tools are often treated as important indicators when determining how to allocate funding. However, because technology transfer offices may often mandate exclusive licensing, the researchers who do seek such licences may bypass open licensing.[170] Some researchers have attempted to found patent pools or pledge patents as a means of encouraging collaborative innovation with patented technologies.[171] Nevertheless, the majority of these patent pools and patent pledges have only found a foothold for information and communications technology patents. For example, only 5 of 178 US patents pledged and listed in the Patent Pledge database operated by American University Law School relate to RNA and DNA sequencing technology.[172]

4.4.4 *Scientific Software Development and the Emerging Knowledge Community*

Knowledge Commons literature has traditionally focused on data sharing. However, the publishing and data sharing environment can be even more complicated for scientific software developers, despite the relative prevalence of standard open source licences.[173] An early foray into the question of open source bioinformatics software by Harvey and McMeekin suggested that initially, bioinformatics developers were often researchers themselves. Therefore

---

[167] Robert English and Charles M. Schweik, 'Identifying Success and Tragedy of FLOSS Commons: A Preliminary Classification of Sourceforge.net Projects' (Paper presented at *First International Workshop on Emerging Trends in FLOSS Research and Development, 2007. FLOSS '07* , May 2007) 5.

[168] Charles M. Schweik and Robert C. English, above n110, 299; Charles Schweik and Robert English, 'Preliminary Steps Toward a General Theory of Internet-Based Collective-Action in Digital Information Commons: Findings From a Study of Open Source Software Projects' (2013) 7(2) *International Journal of the Commons* 234–254 245.

[169] Robert R. Downs et al., 'Community Recommendations for Sustainable Scientific Software' (2015) 3(1) *Journal of Open Research Software*.

[170] Tania Bubela et al., 'Managing Intellectual Property to Promote Pre-Competitive Research: The Mouse as a Model for Constructing a Robust Research Commons' (2012) 22(1) *Journal of Law, Information and Science* 98 120.

[171] Jorge L. Contreras, 'Patent Pledges' (2015) 47(3) *Arizona State Law Journal* 543–608 545.

[172] Jorge Contreras, *Non-SDO Patents and Commitments* (7 September 2017) Program on Information Justice and Intellectual Property <http://www.pijip.org/non-sdo-patent-commitments/>

[173] Jenny Fry, Ralph Schroeder and Matthijs den Besten, 'Open Science in e-Science: Contingency or Policy?' (2009) 65(1) *Journal of Documentation* 6–32 9.

bioinformatics software shared many of the key characteristics of user-produced socio-technical resources.[174] However, using the development of databases and bioinformatics tools as case studies, Harvey and McMeekin argued that gradually, bioinformatics software developers formed their own research communities within molecular biology research.[175] In the context of scientific software development more broadly, Herbsleb and Howison's qualitative research suggests that in the scientific software ecosystem, researchers occupy one of four roles. Firstly, scientist end-users use software to undertake scientific research. Secondly, scientific software developers are responsible for producing and distributing software components. Thirdly, software distribution and execution managers are responsible for distributing software, as well as maintaining scientific computing platforms such as supercomputers. Finally, ecosystem stewards are responsible for the overall functioning of the scientific software ecosystem and its contribution to science. Crucially, Howison and Herbsleb note that contributors may play multiple roles, or indeed switch between roles, within a particular project.[176]

Howison and Herbsleb use these institutional roles to discuss how scientific software is (and should be) created. However, they do not directly address the question of intellectual property rights beyond attribution via citation for software developers. Indeed, a key aspect of the Knowledge Commons literature on open source software has been the focus on social factors underlying development practices. Whilst these are useful indicators of success, ultimately any open source computational biology will not thrive if they lack an assurance of compliance with intellectual property laws. In particular, a lack of compliance may create a bar on cooperation for both domestic and international projects. In addition, returning to the concept of polycentricity, the standard of compliance required may change as the project transitions from a small scale to an international collaboration. Accordingly, the gap that this thesis seeks to fill is the impact of copyright and patent law on the open textualism featured in open source software projects.

4.5 CONCLUSION

This chapter demonstrates how bioinformatics inventions can be protected under copyright law and patent law. In particular, copyright may largely protect the literary aspects of bioinformatics software (namely the source code to implement algorithms for phylogenetic tree construction or sequence alignment). However, copyright protection is limited with

---

[174] Mark Harvey and Andrew McMeekin, 'Public or Private Economies of Knowledge: The Economics of Diffusion and Appropriation of Bioinformatics Tools' (2009) 4(1) *International Journal of the Commons* 481–506 487-8.

[175] Mark Harvey and Andrew McMeekin, above n174, 490.

[176] James Howison et al., 'Understanding the Scientific Software Ecosystem and its Impact: Current and Future Measures' (2015) 24(4) *Research Evaluation* 454–470 455-8.

respect to the functional aspects of bioinformatics algorithms, such as graphical user interfaces and algorithmic operation. With respect to patenting, a bioinformatics algorithm may be patent eligible if it describes something that either renders the algorithm beyond an abstract idea (applying the US standard) or includes a technical effect (applying the European and Australian standard). This consistency is in spite of the recent contraction of patent law in the US to remove computer implemented methods from the scope of patent eligibility Accordingly, this chapter establishes that there is still scope to patent bioinformatics inventions despite attempts to reduce the number of business method patents. This scope for patent eligibility leaves open the question of what impact patent rights, and to a lesser extent copyright protection, have on the success and failure of open source bioinformatics projects.

This chapter also provides a methodological comparison of the predominant frameworks for analysing collaborative research and open socio-technical systems. These frameworks include regulatory capitalism, network theory, social capital theory, and the Knowledge Commons framework. Of these frameworks, the knowledge commons framework represents the culmination of these theories and has been selected for this thesis because of its emphasis on the sustainability of socio-technical resources. That is, without ongoing support, development of an open source bioinformatics project will cease. This chapter concludes by examining the intersection between intellectual property theory and commons governance. In addition, this chapter considers the studies that have applied the Knowledge Commons framework to understand the governance of genomic databases and open source software projects. Between these two areas, this chapter concludes by offering insights into commons-based governance for open source software projects in bioinformatics research.

Chapter 5

QUANTITATIVE RESULTS: A LANDSCAPE OF THE INTERSECTION BETWEEN PATENT RIGHTS AND OPEN SOURCE LICENSING IN BIOINFORMATICS

5.1 INTRODUCTION

This chapter describes a quantitative landscaping model for bioinformatics inventions developed by universities and research institutes. The preceding three chapters have considered the forms of copyright and patent protection which are available to bioinformatics developers. These chapters have also considered what licensing models (including proprietary and open source models) have been discussed within the literature. This chapter first considers what intellectual property rights academic bioinformatics researchers are reliant on. This chapter then considers whether there is a significant overlap between patented and open source bioinformatics. Finally, this chapter considers whether the exercise of patent rights has the potential to impede flow on innovation within bioinformatics research. This quantitative analysis is important because legal and institutional reform into different doctrines of intellectual property law can only be justified through sound empirical evidence. This landscaping model will also inform the qualitative analysis in Chapter Six, which considers whether and why bioinformaticians are relying on these forms of intellectual property rights.

Section 5.2 describes how and why the landscaping methodology employed in this study was developed. It adapts a two-part model for examining university patentee behaviour with respect to bioinformatics software. This model starts with a descriptive landscaping of academic and research institute patent activity by examining patent applications and grants at the United States Patent and Trademark Office (USPTO), the European Patent Office (EPO), the Australian Patent Office (APO) and the IP Office of New Zealand (IPNZO). This landscaping strategy is used to determine what institutes are patenting bioinformatics patents. Further, this landscaping strategy reveals in what areas of bioinformatics research patents are being filed by universities. Section 5.2 finally describes a model for examining whether granted patent rights or open source licensing have an effect on forward citation rates for publications that disclose a bioinformatics algorithm which was later patented (patent publication pairs).

Section 5.3 provides the results of this landscaping exercise, together with an analysis of the results and a discussion of the limitations of this methodology for studying patentee and researcher behaviour. The results reported confirm existing literature and suggest that academics in the US acquire more patents for bioinformatics software than their equivalents in Europe and Australia. Further, academic bioinformatics patent holders prefer to patent in applied research fields. Finally, only a limited sample of patent-publication pairs were able to be identified. However, they nevertheless illustrate that the existence of patent rights themselves was not a sufficient deterrent for ongoing citation and the release of this software under an open source licence is more likely to lead to an increase in forward citations. These results also confirm existing literature. This discussion will frame the qualitative methodology described in Chapter Six, which explores the relationship between formal patent and copyright

laws and user generated norms in bioinformatics communities.

5.2 QUANTITATIVE METHODS AND METHODOLOGY

One of the challenges for this stage of the study involved accurately identifying bioinformatics patents. This identification represents a methodological issue for both researchers studying the effects of bioinformatics software patenting and domain specific researchers who may wish to identify or avoid software patents. In particular, because there is no official definition of a software patent, patent attorneys filing for software patents may use specific drafting strategies. The purpose of this drafting is to send the patent to favourable examiners who would be more likely to grant the patent.[1] This search strategy may result in the dispersion of bioinformatics patents across categories, increasing the difficulty of performing accurate searches. An accurate search requires a balance between type 1 and type 2 errors. In this context, type 1 errors involve failing to collect patents that should be included because of their technological relevance. By contrast, type 2 errors involve retrieving patents that should not be included because of their technological irrelevance.[2] To resolve this problem, this thesis draws upon the considerable literature on mass search strategies to identify software patents to assist in this regard.[3] In particular, three main types of method have been identified to search for bioinformatics patents: classification search strategies; keyword search strategies; and a combination of these two strategies.

5.2.1 *Comparing Classification and Keyword Searching*

Classification search strategies represent a more straightforward means than keyword searching to identify patents. In particular, classification search strategies rely on the classification systems used by different patent offices to direct patents to examiners who have the appropriate scientific qualifications. Classification systems are also used as a means to assist with searching for prior art.[4] The two most widely used classification systems are the International Patent Classification schema (the IPC) and the United States Patent

---

[1] Philip Leith, 'Patenting Programs as Machines' (2007) 4(2) *SCRIPTed: A Journal of Law, Technology and Society* 214–226 224; Florian Berger, Knut Blind and Nikolaus Thumm, 'Filing Behaviour Regarding Essential Patents in Industry Standards' (2012) 41(1) *Research Policy* 216–225 217.

[2] Josh Lerner, 'The New New Financial Thing: The Sources of Innovation Before and After State Street' (Working Paper No 10223, National Bureau of Economic Research, January 2004) 22-3 <http://www.nber.org/papers/w10223>.

[3] These search strategies extend beyond the manual search strategies used by John Allison and Mark Lemley to identify software patents by using claims and descriptions. See (John R. Allison and Mark A. Lemley, 'Who's Patenting What - An Empirical Exploration of Patent Prosecution' (2000) 53(6) *Vanderbilt Law Review* 2099–2174 2110)

[4] Jinseok Park, 'Evolution of Industry Knowledge in the Public Domain: Prior Art Searching for Software Patents' (2005) 2(1) *SCRIPTed: A Journal of Law, Technology and Society* 47–70 64.

Classification schema (USPC).[5] After examining the patents of the six largest software developing companies in the US (based on 1995 revenues),[6] Stuart Graham and David Mowery identified three classes with numerous subclasses for software patents.[7] as described in Table 5.1.

G06F	**Electric Digital Data Processing**
3/	Input arrangements for transferring data to be processed into a form capable of being handled by the computer.
5/	Methods or arrangements for data conversion without changing the order or content of the data handled.
7/	Methods or arrangements for processing data by operating upon the order or content of the data handled.
9/	Arrangements for programme control.
11/	Error detection; Error correction; Monitoring.
12/	Accessing, addressing or allocating within memory systems or architectures.
13/	Interconnection of, or transfer of information or other signals between memories, input/output devices or central processing units.
15/	Digital computers in general.
G06K	Recognition of Data; Presentation of Data; Record Carriers; Handling Record Carriers
9/	Methods or arrangements for reading or recognising printed or written characters or for recognising patterns.
15/	Arrangements for producing a permanent visual presentation of the output data.
H04L	Electric Communication Technique
9/	Arrangements for secret or secure communication.

Table 5.1: Graham and Mowery's IPC classification for software patents in Stuart J.H. Graham and David C. Mowery, 'Intellectual property protection in the U.S. software industry' in Wesley M. Cohen and Stephen A. Merrill (eds), *Patents in the Knowledge Based Economy* (2003), 232

Likewise, Victoria Stodden and Isabel Reich proposed a classification matrix based on the USPC classification for identifying software patents, which is described in Table 5.2 below. Similar classification measures were also used by Stuart Graham and Saurabh Vishnubhakat in examining software patents associated with smartphone litigation.[8]

The advantage of using patent classification marks to search for software patents is that it reduces the potential for an overly inclusive dataset of software patents which may not claim

---

[5] *Strasbourg Agreement Concerning the International Patent Classification*, opened for signature 21st March 1971, 1160 UNTS 483 (entered into force 7th October 1975); Christopher G. Harris, Robert Arens and Padmini Srinivasan, 'Comparison of IPC and USPC Classification Systems in Patent Prior Art Searches' (Paper presented at *Proceedings of the 3rd International Workshop on Patent Information Retrieval* , 2010) 27.

[6] The firms were Microsoft, Adobe, Novell, Autodesk, Intuit and Symantec

[7] Stuart JH Graham and David C. Mowery, 'Intellectual property protection in the US software industry' in Wesley M. Cohen and Stephen A. Merrill (eds.), *Patents in the Knowledge-based Economy* (National Research COuncil 2003)219231,

[8] Isabel Rose Reich and Victoria C. Stodden, 'Software Patents as a Barrier to Scientific Transparency: An Unexpected Consequence of Bayh-Dole' (Paper presented at *The Seventh Annual Conference on Empirical Legal Studies (CELS 2012)* , November 2012) 13; Stuart Graham and Saurabh Vishnubhakat, 'Of Smart Phone Wars and Software Patents' (2013) 27(1) *The Journal of Economic Perspectives* 67–85 75.

PTO Classification Code	Definition
341	Coded Data Generation or Conversion
345	Computer Graphics Processing
370	Multiplex Communications
706	Data Processing: Artificial Intelligence
707	Data Processing: Database and File Management or Data Structures
708	Electrical Computers: Arithmetic Processing and Calculating
716	Computer-aided Design and Analysis of Circuits and Semiconductor Masks
717	Data Processing: Software Development, Installation and Management.

Table 5.2: Isabel Rose Reich and Victoria C. Stodden, 'Software Patents as a Barrier to Scientific Transparency: An Unexpected Consequence of Bayh-Dole' (Paper presented at *The Seventh Annual Conference on Empirical Legal Studies (CELS 2012)*, November 2012), 13.

software as a primary part of the invention.[9] However, the disadvantage of searching by patent classification is under classification. In other words, classification based searching can lead to a failure to identify patents which are software patents but have been filed under other classifications for a myriad of reasons. In addition, patent classifications may change over time to reflect the emergence of new technological fields. These classification changes may potentially undermine attempts to develop a longitudinal study of patenting activity in a particular field. To avoid this problem, James Bessen and Robert Hunt used keyword search strategies to identify the spread of software patents in the US. Specifically, Bessen and Hunt defined computer implemented and software patents as any invention implemented using computer hardware or wholly implemented using software without excluding semiconductor related inventions.[10] This search strategy involves querying the different parts of the patent document. These included the title (which describes the patent), the abstract and specification (which explain exactly how the patent operates, satisfying disclosure requirements) and the claims (which describe the boundaries of the invention).[11] Bessen and Hunt described the following search query as a means of identifying software patents published at the USPTO:[12]

((‘software’ in specification) OR (‘computer’ AND ‘program’ in specification)) AND (utility patent excluding reissues) ANDNOT (‘chip’ OR ‘semiconductor’ OR ‘bus’ OR ‘circuit’ OR ‘circuitry’ in title) ANDNOT (‘antigen’ OR ‘antigenic’ OR

[9] Arti K. Rai, John R. Allison and Bhaven N. Sampat, 'University Software Ownership and Litigation: A First Examination' (2009) 87(5) *North Carolina Law Review* 1519–1570 1529-30.

[10] John R. Allison and Mark A. Lemley, above n3 2110.

[11] James Boyle and Jennifer Jenkins, *Intellectual Property: Law & the Information Society - Cases & Materials: An Open Casebook: 3rd Edition 2016* (CreateSpace Independent Publishing Platform, 3 editioned, 2016) 648.

[12] James Bessen and Robert M. Hunt, 'An Empirical Look at Software Patents' (2007) 16(1) *Journal of Economics & Management Strategy* 157–189 185.

'chromatography' in specification[13]

The advantage of Bessen and Hunt's strategy is accuracy and avoidance of type 1 errors. By using a keyword search bypasses, attempts by patent attorneys to draft software patents so as to seek examination in more favourable examination units are not included in the search results.[14] Bronwyn Hall and Megan MacGarvie note that Bessen and Hunt's search strategy is a particularly accurate means for identifying business method patents. In particular, business method patents are not officially recognised in either the IPC or USPC classification schemas. However, Hall and MacGarvie note that, compared to Graham and Mowrey's class based search system, Bessen and Hunt's search strategy is less accurate for identifying actual software patents. In particular, Bessen and Hunt's strategy leads to more type 2 errors with respect to pure computer hardware patents.[15] MacGarvie and Hall therefore propose a combination search strategy. This combination strategy involved using both the IPC/USPC search query approach Graham and Mowrey apply and the keyword search of title and abstract Bessen and Hunt apply. MacGarvie and Hall then filter their search results to only include patents they identify using both keyword and classification search strategies.[16] MacGarvie and Hall's approach was selected for this thesis as the most appropriate search strategy to reflect the interdisciplinary technological nature of bioinformatics.[17] Accordingly, a combined filtering strategy is discussed in further detail within Section 5.2.2.

5.2.2 *Filtering for Bioinformatics Patents*

The first stage of developing the combined filtering strategy for this thesis involved examining previous literature on bioinformatics patent searching to determine what other strategies had been used and to ensure the search strategy used for this thesis was methodologically sound.[18] Bruce Rasmussen's 2009 study of bioinformatics patenting activity in the US relies upon a

---

[13] Note that the vast majority of patent landscaping studies rely on USPTO data due to the high level of citation information included in that database. See Emily Grant, Megan Van den Hof and E. Richard Gold, 'Patent landscape analysis: A methodology in need of harmonized standards of disclosure' (2014) 39(4) *World Patent Information* 3–10, 7

[14] Ashish Arora, Lee G. Branstetter and Matej Drev, 'Going Soft: How the Rise of Software-Based Innovation Led to the Decline of Japan's IT Industry and the Resurgence of Silicon Valley' (2012) 95(3) *Review of Economics and Statistics* 757–775 759.

[15] Bronwyn H. Hall and Megan MacGarvie, 'The private value of software patents' (2010) 39(7) *Research Policy* 994–1009 999.

[16] Bronwyn H. Hall and Megan MacGarvie, above n15, 998-9.

[17] Francesco Paolo Appio, Antonella Martini and Gualtiero Fantoni, 'The light and shade of knowledge recombination: Insights from a general-purpose technology' (2017) 125(12) *Technological Forecasting and Social Change* 154–165 158.

[18] Emily Grant, Megan Van den Hof and E. Richard Gold, 'Patent Landscape Analysis: A Methodology in Need of Harmonized Standards of Disclosure' (2014) 39(4) *World Patent Information* 3–10 7.

mixed approach. This mixed approach combined USPC classifications and a plain text search of patent descriptions to identify patents that satisfied both the biotechnology and information technology requirements of a bioinformatics patent.[19] These USPC classes are described in Table 5.3 below. Likewise, Neha Mago, Nishad Deshpande and Rajkumar Hirwani used these classes to compare patent activity in bioinformatics research between jurisdictions. Using this landscaping model, Mago, Deshpande and Hirwani were able to conclude that the majority of bioinformatics research occurs in the US and Japan.[20]

PTO Classification Code	**Definition**
382/129	Image Analysis/DNA or RNA pattern reading
702/19	Data Processing: Measuring, Calibrating, or Testing/Biological
702/20	Data Processing: Measuring, Calibrating, or Testing/Gene sequence determination
702/21	Data Processing: Measuring, Calibrating, or Testing/Cell count or shape or size analysis (eg, blood cell)
703/11	Data Processing: Structural Design, Modelling, Simulation and Emulation/Biological or biochemical
703/12	Data Processing: Chemical
703/2	Data Processing: Modelling By Mathematical Expression

Table 5.3: Patent classes for bioinformatics patents identified by Bruce Rasmussen. From Bruce Rasmussen, *Creating and capturing value in the biopharmaceutical sector* (PhD thesis, Victoria University, 2008) 198-9 <http://www.vu.edu.au/research>.

Rasmussen's search strategy manually filters out patents from the initial search results to remove any patents that did not explicitly refer to either information technology or biological inventions as the *primary claim*. Ultimately, Rasmussen search strategy generated a dataset of 364 bioinformatics patents (including academic and commercial patents) issued from 1995 to 2005.[21] However, expanding Rasmussen's search strategy to IPC classes is complicated by the divergence between the IPC classes matching the USPC classes and the IPC classes which are listed as 'bioinformatics classes'. These are the classes which are described in the IPC classes as being most commonly associated with bioinformatics patents.[22] Each of these classes are listed in Tables 5.4 and 5.5 below. Note that the classes in Table 5.5 (which specifically relate to bioinformatics patents) were introduced with the 2010 amendments to the IPC schedule.

---

[19] Bruce Rasmussen, *Creating and capturing value in the biopharmaceutical sector* (PhD Thesis, Victoria University, 2008) 198 <http://www.vu.edu.au/research>.

[20] Neha Mago, Nishad Deshpande and Rajkumar R. Hirwani, 'A Landscape of Bioinformatics Patents - Garnering of IPR in the Field of Bioinformatics' (2017) 51(4) *World Patent Information* 66–78 78.

[21] Bruce Rasmussen, above n19, 199 <http://www.vu.edu.au/research>.

[22] Hyun-Seok Park, 'Preliminary Study of Bioinformatics Patents and Their Classifications Registered in the KIPRIS Database' (2012) 10(4) *Genomics & Informatics* 271–274 272-3.

IPC Code	USPC Code	IPC Definition
G06K 9/00	382/125	Methods or arrangements for reading or recognising printed or written characters for recognising patterns (eg, fingerprints).
G01N 33/48	702/19	Biological material (eg, blood, urine, hemocytometers).
G01N 33/50	702/20	Chemical analysis of biological material
G01N 31/00	702/21	Investigating or analysing non-biological materials by the use of the chemical methods specified in the subgroups
G06G 7/58	703/11	Analog computers for specific processes, systems, or devices, e.g., simulators for chemical processes.
G06G 7/58	703/12	As for G06G 7/58

Table 5.4: Matching USPC and IPC patent classes as conducted by Hyun-Seok Park,'Preliminary Study of Bioinformatics Patents and Their Classifications Registered in the KIPRIS Database' (2012) 10(4) Genomics Informatics 271–274 272-3

A further complication lies in the 'primary' IPC mark associated with a particular patent. A patent may have multiple classes assigned to it for indexing purposes to indicate which technological fields the patent is associated with. In addition to the title, abstract, specification and claims associated with the patent, patent documents also provide the name of the inventor and patent applicant, as well as a measure of the scope of the patent.[23] However, the first listed IPC mark is the dominant measure of the technological and economic field that the patent is intended to cover.[24] Accordingly, including patents that do not include a 'bioinformatics' patent mark as a primary IPC mark may lead to type 2 errors in patent classification.[25] To minimise the risk of type 2 errors within this study, the IPC marks described in Tables 5.4 and 5.5 were used to identify potential bioinformatics patents filed with and granted by the USPTO, the EPO, the Australian Patent Office (APO).[26] Then the patents were manually inspected by the author. Any patents which did not include one of the marks listed in Tables 5.4 and 5.5 as a primary IPC subclass were removed from the data set. The marks in Table 5.4 were used to identify patents that had been filed before 2011 (1 January 2011 being the date the new IPC classification scheme had been introduced) whereas the marks in Table 5.5 were used to identify patents that had been filed after 2011. This search strategy served to balance out the potential

[23] Joshua Lerner, 'The Importance of Patent Scope: An Empirical Analysis' (1994) 25(2) *The RAND Journal of Economics* 319–333 321.

[24] IP Australia, 'Manual of Practice and Procedure' (last update 1 August 2017), s. 2.2.6, 'Responsibility for Furthers, Voluntary Section 104 Amendments'

[25] Henk F. Moed, Wolfgang Glänzel and Ulrich Schmoch, *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (Springer Science & Business Media, 2006) 231.

[26] For the patent lens database, the search heuristic used was classification_ipcr:(G06F191*) OR classification_ipcr:(G06F192*) OR classification_ipcr:(G01N 33/48) OR classification_ipcr:(G01N 33/50) OR classification_ipcr:(G01N 31/00) OR classification_ipcr:(G06G 7/58). The IPC mark G06K 9/00 was not used after a preliminary search revealed too many irrelevant patents.

IPC Classification Code	Definition
G06F 19/00	**Digital computing or data processing equipment or methods, specifically adapted for specific applications (specifically adapted for specific functions G06F17/00; data processing systems or methods specifically adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes G06Q; healthcare informatics G16H)**
G06F 19/10	Bioinformatics, i.e. methods or systems for genetic or protein-related data processing in computational molecular biology (in silico methods of screening virtual chemical libraries; in silico or mathematical methods of creating virtual chemical libraries)
G06F 19/12	For modelling or simulation in systems biology, eg, probablistic or dynamic models, gene-regulatory networks, protein interaction networks or metabolic networks
G06F 19/14	For phylogeny or evolution, e.g., evolutionary conserved regions determination or phylogenetic tree construction
G06F 19/16	For molecular structure, e.g., structure alignment, structural or functional relations, protein folding, domain topologies, drug targeting using structure data, involving two-dimensional or three-dimensional
G06F 19/18	For functional genomics or proteomics e.g., genotype-phenotype associations, linkage disequilibrium, population genetics, binding site identification, mutagenesis, genotyping or genome annotation, protein-protein interactions or protein-nucleic acid interactions
G06F 19/20	For hybridisation or gene expression, eg, microarrays, sequencing by hybridisation, normalisation, profiling, noise correction models, expression ratio estimation, probe design or probe optimisation
G06F 19/22	For sequencing comparison involving nucleotides or amino acids, eg homology search, motif or Single-Nucleotide Polymorphism [SNP] discovery or sequence alignment.
G06F 19/24	For machine learning, data mining or biostatistics, e.g., pattern finding, knowledge discovery, rule extraction, correlation, clustering or classification.
G06F 19/26	For data visualisation, e.g., graphics generation, displays of maps or networks or other visual representations.
G06F 19/28	For programming tools or database systems, e.g. ontologies, heterogenous data integration, data warehousing or computing architectures.

Table 5.5: IPC 'bioinformatics' patent classes. Drawn from World Intellectual Property Organisation (WIPO), *G06F Electrical Digital Data Processing* (01 January 2017) International Patent Classification.

for type 1 errors negatively affecting the accuracy of the database of bioinformatics patents.

5.2.3 *Identifying Academic and Research Institute Bioinformatics Patents*

The next stage of the patent search strategy used in this thesis involved curating the original dataset to only include patents held by academics at universities and research institutes. David Mowery and Bhaven Sampat's study suggests that patents with academic inventors comprise four percent of the patent applications filed with the USPTO.[27] An equivalent study by Franceso Lissoni, Patrick Llerena, Maureen McKelvey and Bulat Sanditov suggests a similar rate for academic patenting at the EPO, with a higher rate for biotechnology patents.[28] Unlike the US, where there is strong evidence of academic patent assignment and ownership, the evidence of European (or Australian and New Zealand) patent ownership sounds significantly more fragmented.[29] Further, previous studies into bioinformatics patenting have not explicitly distinguished between patents filed by research institutes (or joint public-private ventures) and patents filed by private research companies.[30] By contrast, Arti Rai, John Allison and Bhaven Sampat's study of university software patenting in the US examines research universities within the US that had spent the greatest amount of RD expenditure on computer science research over a 20 year period.[31] By focusing on these universities, Rai, Allison and Sampat observed what they described as a 'university effect'; that is, university propensity to patent and litigate software inventions was associated with a greater propensity to acquire software patents.[32]

However, as discussed in further detail in the qualitative methods section in Section 6.2 of Chapter Six, bioinformatics is a more specialised discipline than 'pure' computer science and bioinformatics departments are more likely to be concentrated in particular institutes or attached to particular projects.[33] Accordingly, the patent data set for this thesis was manually

---

[27] David C. Mowery and Bhaven N. Sampat, 'The Bayh-Dole Act of 1980 and University–Industry Technology Transfer: A Model for Other OECD Governments?' (2004) 30(1-2) *The Journal of Technology Transfer* 115–127 120.

[28] Francesco Lissoni et al., 'Academic Patenting in Europe: New Evidence from the KEINS Database' (2008) 17(2) *Research Evaluation* 87–102 94-5.

[29] Chiara Franzoni and Giuseppe Scellato, 'Academic Patenting and the Consequences for Scientific Research' (2011) 44(1) *Australian Economic Review* 95–101 96.

[30] Bruce Rasmussen, above n19 199 <http://www.vu.edu.au/research>; Paul Oldham, Stephen Hall and Geoff Burton, 'Synthetic Biology: Mapping the Scientific Landscape' (2012) 7(4) *PLoS ONE* e34368 5; Saurabh Vishnubhakat and Arti Rai, 'When Biopharma Meets Software: Bioinformatics at the Patent Office' (2015) 29(1) *Harvard Journal of Law & Technology* 206 228-30.

[31] Arti K. Rai, John R. Allison and Bhaven N. Sampat, above n9, 1537-8.

[32] Bronwyn H. Hall and Dietmar Harhoff, 'Recent Research on the Economics of Patents' (Working Paper No 17773, National Bureau of Economic Research, July 2012) 27.

[33] Charles M. Schweik and Robert C. English, *Internet Success: A Study of Open-Source Software Commons* (MIT

curated to identify patents held by universities and research institutes (including both public and private) known to participate in bioinformatics research in the US, the EU, Australia and New Zealand so as to determine the patent activity by these institutions. In particular, this curation strategy fits within a broader narrative in assessing the praxis between releasing software openly and seeking to patent that research software.[34] This curation was conducted by reference to secondary academic literature, including policy publications about computer science and molecular biology technology transfer,[35] as well as institutional technology transfer policy,[36] secondary studies of bioinformatics development and lists of bioinformatics institutes compiled by third parties.[37]

This curation was further refined by examining scientific literature to determine links between spin off companies and particular universities as well as university-industry links to ensure that only relevant patents were identified and included in the dataset.[38] For example, patents filed by the University of Oxford are assigned to ISIS Innovation. ISIS Innovation is a separate technology transfer company established for the purpose of commercialising these inventions.[39] In addition, patents filed by government owned companies that engage in bioinformatics research were also identified and added to the results dataset. These include patents filed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in Australia, the French National Centre for Scientific Research (*Centre national de la recherche scientifique*), the Lawrence Livermore National Laboratory in the US and the VTT Technical Research Centre of Finland. Finally, the sample included patents filed by hospitals,

---

Press, 2012) 171.

[34] Joachim Henkel, 'Selective Revealing in Open Innovation Processes: The Case of Embedded Linux' (2006) 35(7) *Research policy* 953–969 955.

[35] Bruce Rasmussen, above n19, 206 <http://www.vu.edu.au/research>; Annette McLeod, *Returns on Investment: Considerations on Publicly Funded ICT Research and Impact Assessment* (PhD thesis, University of Melbourne, 2016); Jerome H. Reichman, Tom Dedeurwaerdere and Paul F. Uhlir, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge University Press, 2016) 209; Charles E. Cook et al., 'The European Bioinformatics Institute in 2017: Data Coordination and Integration' (2018) 46(D1) *Nucleic Acids Research* D21–D29.

[36] Joaquın M. Azagra-Caro, Luis Plaza-Gómez and Ana Romero-de-Pablos, 'The origin of public research organisation patents: an economic approach' (2007) 16(4) *Research Evaluation* 271–282 272; Tommi Inkinen and Katri Suorsa, 'Intermediaries in Regional Innovation Systems: High-Technology Enterprise Survey from Northern Finland' (2010) 18(2) *European Planning Studies* 169–187; Olof Hallonsten, 'The Third Sector of R&D: Literature Review, Basic Analysis, and Research Agenda' (2017) 35(1) *Prometheus* 21–35 24, 27.

[37] Geoffrey Routh Ph.D, *Institute List* (23 March 2016) Geoff's Bio-Directories <http://www.grouthbio.com/Institute*List.html* >

[38] Rebecca Henderson, Adam B. Jaffe and Manuel Trajtenberg, 'Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965–1988' (1998) 80(1) *The Review of Economics and Statistics* 119–127 186.

[39] Fiona Murray and Scott Stern, 'Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge?: An Empirical Test of the Anti-Commons Hypothesis' (2007) 63(4) *Journal of Economic Behavior & Organization* 648–687 662.

such as the Mount Sinai Medical Centre in the US and the Medical Research Council in the United Kingdom.[40]

5.2.4 *Comparing Patent Publication Pairs*

The final stage of the analysis considers the impact of patent grants on the citation rates of patent-publication pairs in bioinformatics research. Patent publication pairs are scientific publications which are paired with patent applications or patent grants.[41] This analysis goes to the heart of the second question of this thesis.[42] At the outset of the project, there were two key challenges identified with assessing the relationship between patents and open source bioinformatics. The first challenge was that there was no easy means of constructing an 'experimental' and 'control' data set to test the effect of patents.[43] The second challenge was assessing what amounted to a positive or negative impact from patent protection. As Marcus Dapp notes, it can be difficult to assess the negative effect of patents on open source software without identifying specific interactions between patented and open source software developers.[44]

However, one advantage of studying bioinformatics as a field is that the vast majority of bioinformatics software is specialised software developed for scientific research.[45] Accordingly, conventional citation metrics can provide an important, if not entirely complete, picture of the impact of scientific research within a particular field.[46] For the purposes of this thesis, this effect was measured by examining the forward citations of patent publication pairs. The purpose of this analysis was determine whether there was a change in the overall citation

---

[40] Mark Funk, 'Patent Sharing by US Universities: An Examination of University Joint Patenting' (2013) 22(4) *Economics of Innovation and New Technology* 373–391 378.

[41] Francesco Lissoni, Fabio Montobbio and Lorenzo Zirulia, 'Inventorship and Authorship as Attribution Rights: An Enquiry Into the Economics of Scientific Credit' (2013) 95(11) *Journal of Economic Behavior & Organization* 49–69 55; Joshua S. Gans, Fiona E. Murray and Scott Stern, 'Contracting Over the Disclosure of Scientific Knowledge: Intellectual Property and Academic Publication' (2017) 46(4) *Research Policy* 820–835 824.

[42] Note that this model does not consider the impact of copyright laws on open source bioinformatics. This issue will be discussed in more detail in Chapter Six.

[43] James G.S. Wilson, 'On the Value of the Intellectual Commons' in Annabelle Lever (ed.), *New Frontiers in the Philosophy of Intellectual Property* (Cambridge University Press 2012)18122–139,

[44] Marcus M. Dapp, *The Effects of Software Patent Policy on the Motivation and Innovation of Free and Open Source Software Developers* (PhD Thesis, ETH Zurich, 2009) 72.

[45] Charles Schweik and Robert English, 'Preliminary Steps Toward a General Theory of Internet-Based Collective-Action in Digital Information Commons: Findings From a Study of Open Source Software Projects' (2013) 7(2) *International Journal of the Commons* 234–254 241.

[46] Lutz Bornmann and Hans-Dieter Daniel, 'Selecting Scientific Excellence through Committee Peer Review - A Citation Analysis of Publications Previously Published to Approval or Rejection of Post-Doctoral Research Fellowship Applicants' (2006) 68(3) *Scientometrics* 427–440 430-1; Francesco Lissoni, Fabio Montobbio and Lorenzo Zirulia, above n41, 50.

rate for a particular article once the patent associated with that had been granted. If there was a decline in the number of forward citations that a particular article received once a patent had been granted, this decline would suggest some blocking effect. In particular, this result might indicate that other researchers were not directly citing the article due to licensing costs associated with patenting or concerns about patent infringement.[47] In contrast, if there was no statistically significant decline in the number of forward citations after the patent grant, this result would suggest that patents do not have an impact on citation rates.[48] Analysing this effect can be complicated by the fact that raw citation counts are not normally distributed. The absence of a normal or 'Poisson' distribution in citation rates is attributable to differences in the impact factor of individual journals. Further, some articles may have vastly more importance in the scientific literature than others.[49] For example, at least half of all published articles and patents will not be cited, despite the fact that they meet publication requirements or meet the standards for patent eligibility.[50] This effect means that with respect to the forward citation counts for publications, the mean and the variance are not equal, and so the data on forward citations will be over dispersed.[51]

For this reason, it can be difficult to use citation counts as a sole measure of whether patents have an impact on the use of the article or the reported method.[52] In particular, there is significant criticism of the use of citation as a sole measure of scientific impact due to its ease of being manipulated.[53] In the context of bioinformatics software, this effect is demonstrated by Paul Gardner and colleagues. Gardner's findings suggest that academic bioinformatics software developers will develop either faster or more accurate algorithms to publish in higher ranked journals.[54] Accordingly, the first step for the paired analysis approach used in this

---

[47] Tom Magerman, Bart Van Looy and Koenraad Debackere, 'Does Involvement in Patenting Jeopardize One's Academic Footprint? An Analysis of Patent-Paper Pairs in Biotechnology' (2015) 44(9) *Research Policy* 1702–1713 1709.

[48] Clarisa Long, 'Patent Signals' (2002) 69(2) *The University of Chicago Law Review* 625–679 647; Pierre Azoulay, Waverly Ding and Toby Stuart, 'The Impact of Academic Patenting on the Rate, Quality and Direction of (Public) Research Output' (2009) 57(4) *The Journal of Industrial Economics* 637–676 663.

[49] Jeffrey L. Furman, Kyle Jensen and Fiona Murray, 'Governing Knowledge in the Scientific Community: Exploring the Role of Retractions in Biomedicine' (2012) 41(2) *Research Policy* 276–290 285.

[50] David Popp, 'They Don't Invent Them Like They Used to: An Examination of Energy Patent Citations Over Time' (2006) 15(8) *Economics of Innovation and New Technology* 753–776 757.

[51] Jerry Hausman, Bronwyn H. Hall and Zvi Griliches, 'Econometric Models for Count Data with an Application to the Patents-R & D Relationship' (1984) 52(4) *Econometrica* 909–938 921-2.

[52] Michael Callaham, Robert L. Wears and Ellen Weber, 'Journal Prestige, Publication Bias, and Other Characteristics Associated With Citation of Published Studies in Peer-Reviewed Journals' (2002) 287(21) *JAMA* 2847–2850 2848-9.

[53] James Howison and Julia Bullard, 'Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature' (2015) *Journal of the Association for Information Science and Technology* 2.

[54] Paul Gardner et al., 'A Meta-Analysis of Bioinformatics Software Benchmarks Reveals that Publication-Bias

thesis was to identify publications associated with patents. This pairing was achieved by examining the citations included with each patent. In the alternative, pairing was achieved by examining the publication record of the patent applicant to determine whether the publication contained the algorithm disclosed within the patent. This search strategy follows that used by Fiona Murray and Scott Stern to identify biotechnology patent publication pairs published in *Nature* between 1997 and 1999.[55] From this search strategy, Murray and Stern built a dataset of 340 unique articles, with 169 of those articles associated with a patent at October 2003.[56] That sampling strategy allowed Murray and Stern to use a 'control group' of non patented articles and an experimental group of patented articles to compare the change in citation rates. By contrast, the though the sampling method used for this thesis did not allow for a control group. Nevertheless, it was still possible to observe the effect of an exogenous shock (namely the granting of the patent) on forward citation rates before and after the acquisition of a patent associated with the publication.[57]

Negative binomial regression modelling was used to assess the change in the citation rate for bioinformatics articles which disclosed a novel bioinformatics algorithm that was later the subject of an academic patent in the US, the EU or Australia, using the list of patents that had been compiled using the patent landscaping strategy described in Sections 5.2.1 to 5.2.3. Based on an analysis of the data, there were a relatively low number of zero counts, so a zero inflated Poisson distribution model was not used for this project.[58] These articles were identified by either reviewing the full text for each patent and then identifying the matching algorithm, by either searching the author's publication record or searching the citations listed in the patent for articles authored by the patent applicant.[59] Negative binomial regression was used for the

---

Unduly Influences Software Accuracy' (2016) *bioRxiv* 092205 4-5.

[55] Fiona Murray and Scott Stern, above n39, 661.

[56] Fiona Murray and Scott Stern, above n39 661.

[57] An exogenous shock is a term used in econometrics research to describe a change in the normal economic conditions within a particular market place. Within the scientific publication space, the normal economic conditions that will determine citation rates include the impact factor of the journal as well as the scientific importance of the results disclosed within that paper. The acquisition of a patent was introduced as a control factor to constitute an exogenous shock. In other words, the patent will influence whether the publication is more or less highly cited. (William H. Greene, *Econometric Analysis* (Pearson/Prentice Hall, 8th editioned, 2012) 259). As explained below, using negative binomial regression with panel data allows the use of dummy variables to assess the effects of a change in policy before and after a particular incident (in this case, the acquisition of a patent). Although control groups can be used for negative binomial regression on count data, control variables are primarily used to determine the impact of different factors on a set of count data .

[58] Achim Zeileis, Christian Kleiber and Simon Jackman, 'Regression Models for Count Data in R' (2008) 27(8) *Journal of Statistical Software* 1–25 8.

[59] Note that the author has an undergraduate degree in computer science, and was able to identify these algorithms through an in-depth textual analysis of each article (Fiona Murray, 'Innovation as Co-Evolution of Scientific and Technological Networks: Exploring Tissue Engineering' (2002) 31(8) *Research Policy* 1389–1403 1392-3). Also note that due to limitations on the 'grace period' for patent applications, publications were matched depending on whether they were published at the same time as the patent was filed

present study because it is appropriate for overly-dispersed count variables where the dependent variable is not equal to the mean. An example of such data are article and patent citations, where certain articles may have a higher impact than others.[60]

Citation data relevant to each of the articles under consideration in the present study was then collected, starting from the first year after the articles were published to the last year of complete annual citation data (which for all articles was 2017).[61] Accordingly, the number of annual citations became the dependent variable. Years since publication was included as an independent variable in this study to determine whether the citation count for articles increased over time or whether the citation count decayed as time since the article was published elapsed.[62] This citation data was collected using Scopus and cross checked using Thompson Reuters Web of Science for validity.[63] In addition, self-citations were removed from the final counts, as these citations would presumably not incur any licensing fees or demonstrate any evidence of a downstream effect on citation rates.[64]

To assess the impact of the patent grant on citation rates, a number of dummy variables were introduced as control variables. These dummy variables were as follows:

Time: This variable was a fixed variable indicating the number of years that had elapsed since the article was published. There is equivocal evidence as to whether citation rates decay over time before and after an article is published.[65] As such, the time since publication was used as a control variable.

Country: This dummy variable indicated where the patent applicant (represented as an institution) was located and was set to one of three variables; that is, US, EU (for European Union member states or the United Kingdom) or AU (which includes Australia and New Zealand). This dummy variable was included in the present study to determine whether geographic concentration leads to an increase or decrease in the flow of scientific knowledge

---

[60] Jeffrey L. Furman and Scott Stern, 'Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research' (2011) 101(5) *American Economic Review* 1933–1963 1947.

[61] Hendrik P. van Dalen and Kéne Henkens, 'Signals in Science - On the Importance of Signaling in Gaining Attention in Science' (2005) 64(2) *Scientometrics* 209–233 221.

[62] Neil C. Thompson, Arvids A. Ziedonis and David C. Mowery, 'University Licensing and the Flow of Scientific Knowledge' (2018) 47(6) *Research Policy* 1060–1069 1063.

[63] Both of these sources are peer reviewed citation sources that provide the most accurate data about the impact factor for individual articles as well as journal series (Heather A. Piwowar and Todd J. Vision, 'Data Reuse and the Open Data Citation Advantage' (2013) [1] *PeerJ* e175 6).

[64] Compare with citation rates in scientific fields that do not involve commercialisable technology and where there is unlikely to be a change in citation rates (Glenn D. Walters, 'Predicting Subsequent Citations to Articles Published in Twelve Crime-Psychology Journals: Author Impact versus Journal Impact' (2006) 69(3) *Scientometrics* 499–510 502).

[65] Tom Magerman, Bart Van Looy and Koenraad Debackere, above n47, 1708.

as expressed through citation rates.[66]

Journal Impact Factor: This variable was used to measure the number of citations that all articles published in a particular journal has relative to the number of articles published in that journal during the same period of time.[67] Whilst Murray and Stern exclusively sampled patent publication pairs from the journal *Nature* (so that each article would have an identical impact factor), Tom Magerman, Bart Van Looy and Koenraad Debackere controlled for journal impact on the grounds that journal impact is representative of the general quality of the journal.[68] Because the patent publication pairs in this study were distributed across a number of different journals, journal impact factor was introduced as a control variable.

Private Partner: This variable was a dummy variable used to represent whether there was a private entity (either associated with or separate to the university or research institute applying for the patent) listed as a joint patent applicant with the research institute. This factor is important to control for, as university-industry collaboration may be indicative of the utility of basic scientific research.[69] In addition, the fact that a university inventor has co-applied or assigned their patent to a private partner could indicate that they have adopted a backdoor route to commercialising a faculty invention, and that the private partner may be willing to enforce that patent against infringers.[70]

USPTO Patent Grant: This dummy variable was used in the present study to indicate whether the USPTO had awarded a patent for the algorithm disclosed within the article. A value of 0 indicated that a patent had not yet been awarded for the algorithm by the USPTO, whereas a value of 1 indicated that the patent had been granted and was in force.

EPO Patent Grant: This dummy variable was used in the present study to indicate whether the EPO had awarded a patent for the algorithm disclosed within the article. A value of 0 indicated that a patent had not yet been awarded for the algorithm by the EPO, whereas a value of 1 indicated that the patent had been granted and was in force.

APO Patent Grant: This dummy variable was used in the present study to indicate whether the APO had awarded a patent for the algorithm disclosed within the article. A value of 0 indicated that a patent had not yet been awarded for the algorithm by the APO, whereas a value of 1 indicated that the patent had been granted and was in force.

---

[66] Adam B. Jaffe and Manuel Trajtenberg, 'Flows of Knowledge from Universities and Federal Laboratories: Modeling the Flow of Patent Citations over Time and Across Institutional and Geographic Boundaries' (1996) 93(23) *Proceedings of the National Academy of Sciences* 12671–12677 12672.

[67] Eugene Garfield, 'The History and Meaning of the Journal Impact Factor' (2006) 295(1) *Journal of the American Medical Association* 90–93.

[68] Tom Magerman, Bart Van Looy and Koenraad Debackere, above n47, 1706.

[69] Negin Salimi, Rudi Bekkers and Koen Frenken, 'Does Working with Industry Come at a Price? A Study of Doctoral Candidates' Performance in Collaborative vs. Non-Collaborative Ph.D. Projects' (2015) 41-42(4) *Technovation* 51–61 51-2.

[70] Jerry Thursby, Anne W. Fuller and Marie Thursby, 'US faculty patenting: Inside and outside the university' (2009) 38(1) *Research Policy* 14–25 18.

Open Source: This dummy variable was used in the present study to indicate whether the algorithm was included in the article as an open source release when published. Because none of the identified articles described an open source software package that was released under a restrictive open source licence, a distinction was not drawn between different types of open source licences (permissive or restrictive).

The two hypotheses for this analysis were expressed as follows:

Hypothesis 1: The grant of a patent before the USPTO, the EPO or the APO will have a negative effect on forward citations for a matched publication pair.

Hypothesis 2: A publication that discloses an algorithm under an open source licence that is later patented will receive more citations than an article that does not release an algorithm under an open source licence.

5.3 RESULTS

5.3.1 *Initial Patent Landscaping*

Using the search heuristic strategy described above, an initial patent landscaping model was used to identify patents in the US, the EU and Australia. The purpose of this landscaping strategy for this thesis was to develop an exploratory model of trends within bioinformatics patenting.[71] This basic landscaping model is useful for examining technological change (such as through comparing filing and grant rates within different patent classes and subclasses), as well as the number of patents filed by different institutions. In the present study, these patents were first filtered to include those published from the 1st of January 2006 to the 1st of January 2017 (to create a date range between 2006 and 2017). This date range was used because 2006 represented the first year when Next Generation Sequencing techniques were publicised.[72] These application and grant statistics are shown in Figures 5.1, 5.2 and 5.3 with a comparison of US, EU and Australian patent filing rates for university filed inventions. In addition, the distribution of patent documents by IPC classes is described in Tables 5.6, 5.7 and 5.8.

---

[71] Bruce Rasmussen, above n19, 199 <http://www.vu.edu.au/research>.

[72] Wilhelm J. Ansorge, 'Next-generation DNA sequencing techniques' (2009) 25(4) *New biotechnology* 195–203 317.

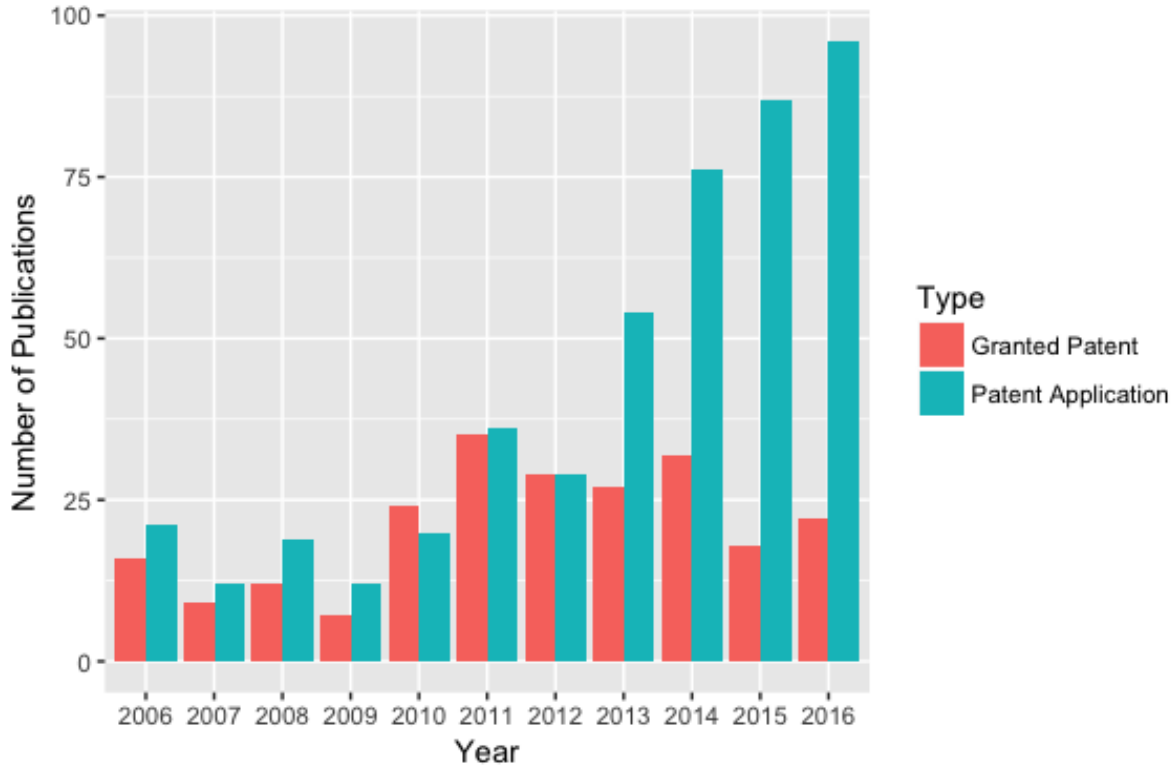

Figure 5.1: Patent applications and grants filed at the USPTO with a bioinformatics primary IPC mark, published between 2006 and 2016 identified in the present study. Data were collected from the Patent Lens dataset hosted by CAMBIA and the Queensland University of Technology (QUT) and was cross referenced using the USPTO Patent Public Pair. Data was then cleaned to remove any non-university software patents classes.

IPC Code	**Number of Documents**
G06F 19/10	9
G06F 19/12	94
G06F 19/14	6
G06F 19/16	66
G06F 19/18	95
G06F 19/20	48
G06F 19/22	98
G06F 19/24	60
G06F 19/26	10
G06F 19/28	20
G06G 7/58	56
G01N 33/48	124
G01N 33/50	11
Total	697

Table 5.6: Bioinformatics patent applications and grants filed by research institutes at the USPTO and published between 2006 and 2016, divided by primary IPC mark, identified in the present study. Primary IPC mark was collected by examining patent documents through the USPTO Public Pair database.

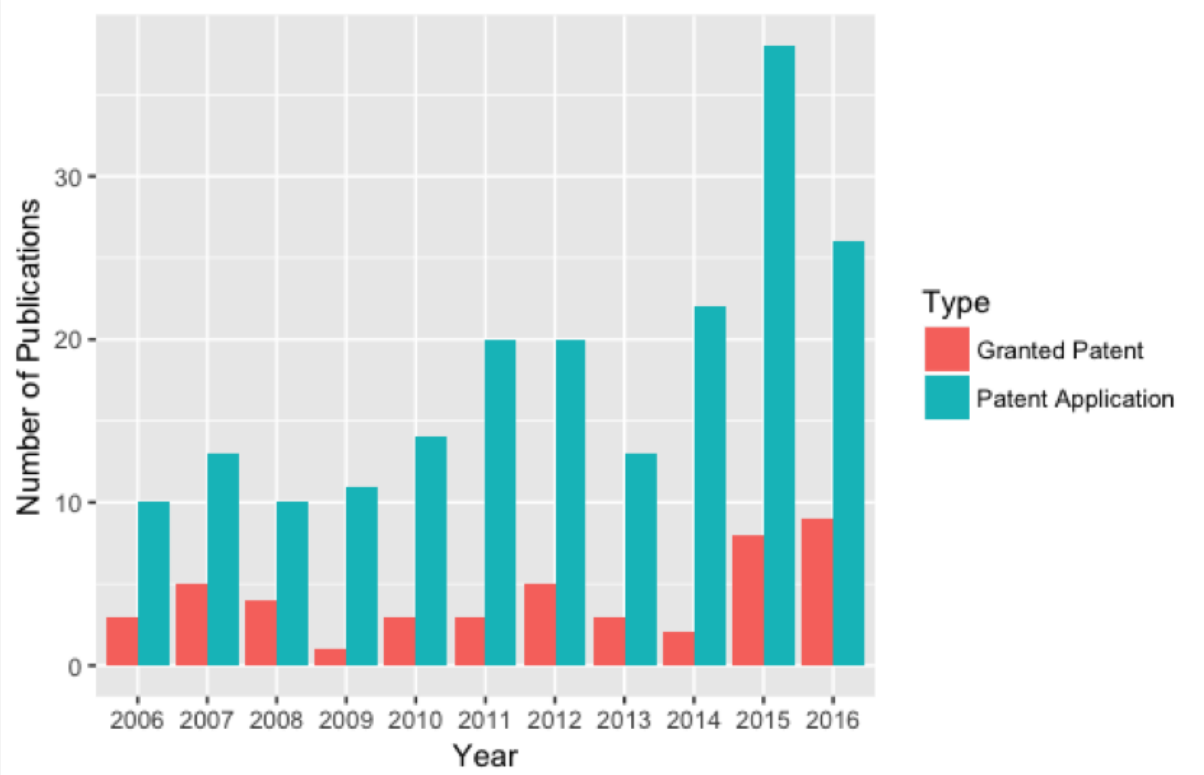

Figure 5.2: Patent applications and grants filed at the EPO with a bioinformatics primary IPC mark, published between 2006 and 2016 identified in the present study. Data were collected from the Patent Lens dataset hosted by CAMBIA and the Queensland University of Technology (QUT) and was cross referenced using the EPO P@tentScope database. Data was then cleaned to remove any non-university software patents classes.

IPC Code	**Number of Documents**
G06F 19/10	20
G06F 19/12	20
G06F 19/14	5
G06F 19/16	21
G06F 19/18	32
G06F 19/20	20
G06F 19/22	33
G06F 19/24	20
G06F 19/26	3
G06F 19/28	5
G06G 7/58	5
G01N 33/48	47
G01N 33/50	12
Total	243

Table 5.7: Patent documents filed by research institutes at the EPO and published between 2006 and 2016, divided by primary IPC mark, identified in the present study. Primary IPC mark was collected by examining patent documents through the EPO Esp@cenet database.

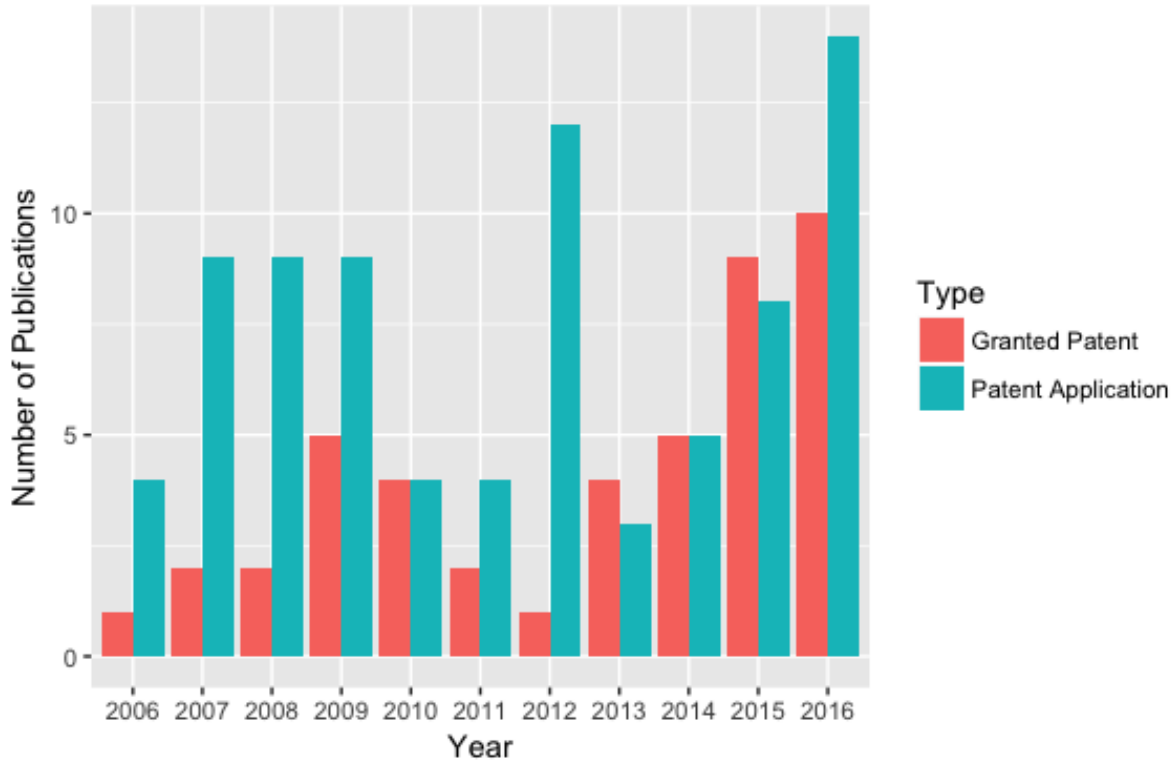

Figure 5.3: Patent applications and grants filed at the APO with a bioinformatics primary IPC mark, published between 2006 and 2016 identified in the present study. Data were collected from the Patent Lens dataset hosted by CAMBIA and the Queensland University of Technology (QUT) and was cross referenced using the APO AusPat database. Data was then cleaned to remove any non-university software patents classes.

IPC Code	**Number of Documents**
G06F 19/10	29
G06F 19/12	5
G06F 19/14	0
G06F 19/16	2
G06F 19/18	3
G06F 19/20	3
G06F 19/22	2
G06F 19/24	2
G06F 19/26	5
G06F 19/28	0
G01N 33/48	68
G01N 33/50	8
G06G 7/58	1
Total	125

Table 5.8: Patent documents filed by research institutes at the APO, divided by primary IPC mark, identified in the present study. Note that no bioinformatics patents were successfully granted at the NZIPO.

Each of the histograms (shown in figures 5.1 to 5.3) demonstrate a gradual increase in the number of bioinformatics patents starting from 2006. The only noticeable change in this trend is a decrease in the number of patent publications after 2008 which took approximately three years to correct. However, there is a steady trend towards an increasing number of patent applications and patent grants over the period of study. This trend could be attributable to a number of factors. These factors may not necessarily be discerned from a pure quantitative analysis of patent filing and grant rates. One possible explanation, as postulated by Samuel Kortum and Josh Lerner, is that a change in management practices and an increased emphasis on patent protection and formal technology transfer leads to an upwards trend in patent activity, irrespective of technological field or jurisdictional changes.[73] However, this increase may also be attributed to IPC class changes in bioinformatics that complicate such longitudinal analysis by introducing new patent classes for patent attorneys to file under, as demonstrated by Tables 5.6 to 5.8.

In addition, Figures 5.2 and 5.3 demonstrate that there is a smaller ratio of patent grants to patent applications at the EPO and the APO relative to the USPTO. What these descriptive statistics suggest is that potential academic bioinformatics patent applicants are concentrating on filing their patents with the USPTO as opposed to the EPO and the APO. These results might suggest that academic bioinformaticians who wish to commercialise their software are more likely to file at the USPTO as opposed to the EPO or the APO. A possible explanation for this tendency is that identified by Rai, Allison and Sampat in that universities and research institutes which have a tendency towards patenting activity across *all* technological fields are more likely to engage in software patenting. This trend can be compared to universities which do not engage in heavy patenting and are therefore less likely to patent. Another possible explanation is that these universities that patent more frequently target larger markets, such as the US, as opposed to smaller markets, such as Australia and New Zealand. A breakdown of patent filing rates amongst the top 10 patent applicants identified in this patent landscaping model for this study are shown in Tables 5.9, 5.10 and 5.11 below.

As these tables demonstrate, the University of California (and spinoff firms Five3Genomics LLC, Genomatica Inc, GT Life Sciences Inc and Nantomics LLC) are responsible for the vast majority of patents filed across each of the three jurisdictions. In addition, the Massachusetts Institute of Technology is the sixth, eighth and seventh highest ranking patenting institution across all three jurisdictions. Finally, the University of Washington ranked 2nd for patent documents at both the EPO and the APO (as well as having filed 9 documents at the USPTO), and although Stanford University did have a relatively high

[73] Samuel Kortum and Josh Lerner, 'What is Behind the Recent Surge in Patenting?' (1999) 28(1) *Research Policy* 1–22 289-90.

Institution	Number of Documents
University of California (including subsidiaries)	65
New York University	32
Stanford University	30
California Institute of Technology	24
University of Texas	16
Massachusetts Institute of Technology	15
BGI Genomics Ltd	14
Columbia University	13
Wisconsin Alumni Research Foundation	13
Harvard University	12

Table 5.9: Top 10 patent filing bioinformatics institutions at the USPTO between 2006 and 2016, in ascending order.

Institution	Number of Documents
University of California (including subsidiaries)	24
University of Washington	11
University of Cambridge	11
Chinese University of Hong Kong	9
Institute of Systems Biology	8
Max Planck Society	8
University of North Carolina	7
Massachusetts Institute of Technology	5
Sloan Kettering Institute of Cancer	5
University of Oxford	4

Table 5.10: Top 10 patent filing bioinformatics institutions at the EPO between 2006 and 2016 in ascending order.

Institution	Number of Documents
University of California (including subsidiaries)	48
University of Washington	11
CSIRO	7
Stanford University	6
State University of New York	6
Fred Hutchinson Cancer Research Center	5
Massachusetts Institute of Technology	5
University of North Carolina	5
University of Texas	4
Baylor Research Institute	4

Table 5.11: Top 10 patent filing bioinformatics institutions at the APO between 2006 and 2016 in ascending order.

number of bioinformatics patents filed at the USPTO and the APO, it ranked comparatively low at the EPO, with only one patent filed. All of these research institutes are noted for having highly developed pathways for technology transfer. This finding is consistent with findings in other studies, even where technology transfer is not directly focussed on exploitation of technology.[74] Therefore, the findings of the present study would seemingly confirm that

[74] Donald S Siegel, David Waldman and Albert Link, 'Assessing the Impact of Organizational Practices on the

universities with highly developed technology transfer offices are more likely to engage in patenting.[75] Another possible explanation is that life science researchers are clustered in particular locations, such as California, Washington and Boston in the US.[76] In particular, Rasmussen's study suggests that geographic proximity can be a key indicator of propensity to patent.[77]

This regional filing trend is reflected in the other two jurisdictions. For example, the University of Cambridge, the Max Planck Society and the University of Oxford are the fourth, sixth and tenth most frequently listed patent applicants at the EPO. Likewise, the CSIRO is the third most frequently listed patent applicant before the APO. The high ranking of the first three applicants as identified in this study would appear to contradict what has been dubbed the 'European Academic Paradox'.[78] Accordingly to this theory, European research institutes conduct strong fundamental scientific research but struggle to convert this research into applied technology.[79] However, by contrast European academic institutes are only marginally less active in patenting bioinformatics algorithms as their US equivalents. Further, as David Audretsch and Devrim Göktepe-Hultén suggest, the absence of university software patents in Europe can be more accurately described as an absence of 'university *owned* patents'. That is, a significant amount of university research is diffused into private industry, either through industry work or through industry consultation. Any resultant patents are then transformed into privately owned patented technology, or 'university invented patents'.[80] Accordingly, as part of the patent landscaping model used in this thesis, spin off firms were identified for privately filed patents developed using university research. This additional layer of

---

Relative Productivity of University Technology Transfer Offices: an Exploratory Study' (2003) 32(1) *Research Policy* 27–48 44; Katharine Ku and James Henderson, 'The MTA—Rip it Up and Start Again?' (2007) 25(7) *Nature Biotechnology* 721–721 721; Andrew J. Nelson, 'From the Ivory Tower to the Startup Garage: Organizational Context and Commercialization Processes' (2014) 43(7) *Research Policy* 1144–1156 1148-9.

[75] Yonghong Wu, Eric W. Welch and Wan-Ling Huang, 'Commercialization of University Inventions: Individual and Institutional Factors Affecting Licensing of University Patents' (2015) 36-37(February-March) *Technovation* 12–25 22-3.

[76] Philip Cooke, 'Life Sciences Clusters and Regional Science Policy' (2004) 41(5-6) *Urban Studies* 1113–1131 1116; Philip Cooke, 'Biosciences and the Rise of Regional Science Policy' (2004) 31(3) *Science and Public Policy* 185–197 187.

[77] Bruce Rasmussen, above n19, 206 <http://www.vu.edu.au/research>.

[78] Robert J. W Tijssen and Erik van Wijk, 'In Search of the European Paradox: an International Comparison of Europe's Scientific Performance and Knowledge Flows in Information and Communication Technologies Research' (1999) 28(5) *Research Policy* 519–543 520.

[79] Christian O. Fisch et al., 'University patenting: a comparison of 300 leading universities worldwide' (2015) 40(2) *The Journal of Technology Transfer* 318–345 336-7.

[80] David B. Audretsch and Devrim Göktepe-Hultén, 'University Patenting in Europe: Does Faculty Ownership of Intellectual Property Impede University Technology Transfer?' in Albert N. Link, Donald S. Siegel and Mike Wright (eds.), *The Chicago Handbook of University Technology Transfer and Academic Entrepreneurship* (University of Chicago Press 2015) 199-205.

bibliometric analysis has been used in other studies to identify privately owned patents.[81] For example, there is at least one example of a bioinformatics firm, Lion Biosciences, being developed as a private spinoff from the European Molecular Biology Laboratory (EMBL).[82] The qualitative model described in Chapter Six is used to ascertain more details about technology transfer for bioinformatics inventions at European research institutes. For Australia, CSIRO is the highest ranked local patent applicant before the APO. This finding is indicative of the nature of the Australian innovation system, which is discussed in further detail in Chapter Seven.

Information about bioinformatics patents can also be gleaned from the document categories under which various patents are filed. For both the USPTO and the EPO a very small number of patents were identified as being filed under the class G06F1914 (which pertains to algorithms for evolutionary tree construction and phylogenetics research). Yet at the APO there were no patents identified as being granted for patents under this class between 2006 and 2016. This finding may be attributable to the nature of phylogenetics software, which is is frequently used in evolutionary biology and ecology research. It is possible that these fields do not encourage significant commercialisation of the software. This finding is supported by a specific examination of a patent filed by the American Museum of Natural History.[83] Although the USPTO granted the patent in May 2006, the patent application lapsed before the APO in May 2005.[84] Likewise, there are only 20 and 5 patent documents respectively published at the USPTO and the EPO (at 0 at the APO) with respect to class G06F1928. As discussed above relates to programming tools for database systems, ontologies, heterogenous data integration and data warehousing. As for phylogenetics software, a possible explanation for this disparity is the fact that database development is already populated by commercially useful open source solutions.[85] In these circumstances, the number of software algorithms that could meet the requisite standards of novelty would be expected to be relatively small. However, an in-depth examination of certain patents could reveal another potential explanation. In particular, one patent granted to the University of California by the USPTO describes a patent for high throughput sequence analysis.[86] In particular, paragraph 1 of the background of the invention describes the challenges of computing 'significant levels'

---

[81] Peter van Dongen, Hester Tak and Eric Claassen, 'Policies and patenting to stimulate the biotechnology sector: evidence from The Netherlands' (2018) 46(1) *Science and Public Policy* 136–147 4.

[82] Philip Cooke, 'European Asymmetries: A Comparative Analysis of German and UK Biotechnology Clusters' (2007) 34(7) *Science and Public Policy* 454–474 458.

[83] US7043371B2, *Method for Search Based Character Optimization*

[84] AU2003259899, *Method for Search Based Character Optimization*

[85] See also (Paul B. de Laat, 'Copyright or copyleft?: An analysis of property regimes for software development' (2005) 34(10) *Research Policy* 1511–1532 1523)

[86] US 9646134B2, Bambam: parallel comparative analysis of high-throughput sequencing data

of genomic data. This description may suggest that machine learning algorithms, having become increasingly important in bioinformatics, will be increasingly patented.

This conclusion is partially supported by the remaining patented classes. For example, patents belonging to class G06F1924 constitute a minority of all patent documents (60 documents before the USPTO, 20 before the EPO and 2 before the APO). However, there has a steady increase in the number of patent documents published in this class. For example, 14 patent documents were published in 2016, compared to 9 in 2015, before the USPTO. This gradual increase also reflects the increasing importance that universities place on the role of machine learning algorithms in bioinformatics. Further, compared to classes G06F1914 and G06F1928, considerably more patents are filed in the classes G06F1912 and G06F1922. These classes concern software for systems biology research and sequencing software respectively. Vandamme notes that systems biology software and sequencing software may play a considerable role in improving efficacy in drug discovery and drug development. Accordingly, it is possible that university inventors may see that there is potential for successful technology transfer for these software tools.[87] Finally, the present patent landscaping model identified that the two classes G01N3348 and G01N3350 feature in a high number of patent documents. In turn, the presence of these documents which may be indicative of a trend towards filing patents for combined software and non software inventions. The purpose of this combined filing could be to receive more favourable examination and take advantage of leniency towards embedded software patents, particularly before the EPO. This finding would be consistent with the discussion of patentable subject matter requirements in Sections 3.3.1 to 3.3.3 of Chapter Three, which demonstrated that the European Patent Board of Appeals's interpretation of Article 52(2) favoured patents that provided a technical effect.

In addition, the IPC class descriptions for G01N3348 and G01N3350 do not explicitly state that they are 'algorithm' or data processing classes. These descriptions may indicate that patent applicants or technology transfer offices are constructing their applications for assignments in these particular classes. One explanation for this phenomenon might be their patents describe a novel way of using a pre-existing general purpose algorithm that alone would not meet the threshold for patent eligibility. Another possible explanation is that patent examiners are targeting established types until more consistent guidelines on bioinformatics patent eligibility emerge.[88] Finally, it is worth noting that the count statistics identified in the present study do not discriminate between patent applications and patent grants. In other words, patent applications may be drafted speculatively to cover as much of the invented

---

[87] Drieke Vandamme et al., 'Systems Biology-Embedded Target Validation: Improving Efficacy in Drug Discovery' (2014) 6(1) *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 1–11 3.

[88] Trent Ostler and Michael Gollin, 'Which Types of Bioinformatics Inventions are Eligible for Patent Protection?' (2015) 21(2) *Journal of Commercial Biotechnology* 76–82 79.

technology as possible. Whilst a broader initial patent application may be preferable, the eventual grant may be significantly more narrow.[89] However, the present patent landscaping model used in this thesis cannot ultimately explain the rationale amongst inventors or technology transfer officers for filing patents under particular patent subclasses. Chapter Six explores these issues using a qualitative approach. For now, the next section of this Chapter turns to address the relationship between the patent publication pairs identified as part of this research.

5.3.2 *Patent Publication Pairs*

The filtering method described in section 5.2.4 revealed a data set of 86 publications that were paired with bioinformatics patents granted in the US, the EU and Australia.[90] The citation counts collected for each article ranged from 7 years after publication to 20 years after publication, for a total of 1134 annual citation counts. These statistics were loaded into a common separated variable (CSV) file, which was then uploaded into R, a computational statistical analysis platform.[91] A negative binomial regression was conducted using the `glm.nb` command from the `MASS` library in R.[92] The results of this negative binomial regression are reported in the command sequence below.

```
   library(MASS)
   pp.nb.model <- glm.nb(formula = no.citations ~time + country
+ journal.impact + private.partner + us.patent.granted +
eu.patent.granted + au.patent.granted + open.source, data =
neg.binomial.citation.rates)
   Call:
   glm.nb(formula = no.citations ~time + country +
journal.impact + private.partner + us.patent.granted +
eu.patent.granted + au.patent.granted + open.source, data =
neg.binomial.citation.rates, init.theta = 0.7429424119, link =
log)
   Deviance Residuals:
```

---

[89] Paul Oldham and Anthony Mark Cutter, 'Mapping Global Status and Trends in Patent Activity for Biological and Genetic Material' (2006) 2(2) *Genomics, Society and Policy* 62 67.

[90] Although there has been some discussion of the scope of copyright and patent law in New Zealand, and there were a number of New Zealand interviewees whose perspectives were described in the results in Chapter Six, there were no patents and patent publication pairs identified as being filed at the NZIPO. Chapter Six will discuss this issue in further detail.

[91] See (R Core Team, *R: A Language and Environment for Statistical Computing*. (2013) R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org/>)

[92] For further details see (James E. Monogan III, *Political Analysis Using R* (Springer International Publishing, 2015) 119-121)

```
 Min 1Q Median 3Q Max
-2.528 -1.059 -0.513 0.059 4.536
```

The coefficients are listed in table 5.12.

Table 5.12: Negative Binomial Regression

Parameter	Estimate	Std. error	Pr(> |z|)
(Intercept)	2.2346	0.1790	< 0.0000 ***
Time Since Publication	-0.0242	0.0106	0.0222
Researcher from EU	-0.9243	0.2073	0.000 ***
Researcher from US	0.1332	0.1770	0.4516
Journal Impact	0.0618	0.0041	0.0000 ***
Private Partner	-0.7346	0.1133	0.0000 ***
US Patent Granted	0.1374	0.0976	0.1589
EU Patent Granted	0.2741	0.1606	0.0879 .
AU Patent Granted	-0.0438	0.1092	0.6886
Open Source	-2.7315	0.5429	0.000 ***
Not Open Source	0.9172	0.0785	0.000 ***

Negative Binomial Model, Determinants of Forward Citations for articles that were subsequently patented at the USPTO, the EPO and the APO. Statistical significance in parentheses *** denotes p<0.001, ** denotes <0.01 and * denotes p<0.05. Total observations = 1134.

To test this model, a Poisson model was also calculated, and the residuals from this model were calculated with a five percent critical value chi-squared test. The deviance of these residuals was then compared to the residual for the negative binomial regression model.[93]

```
poisson.model <- glm(formula = no.citations ~time + country
+ journal.impact + private.partner + us.patent.granted +
eu.patent.granted + au.patent.granted + open.source, data =
neg.binomial.citation.rates)
> qchisq(0.95, df.residual(poisson.model))
1202.073
> deviance(nbr.model)
1299.259
```

Before turning to the results obtained from these analyses, it is important to note that there are a number of limitations to these results. Firstly, due to differences in academic publication impact factor and citation numbers, the data on citation is highly skewed in that certain articles will be cited more extensively than other articles. Secondly, forward citations for academic publications do not necessarily indicate the only way that a publication, or an algorithm

[93] Germán Rodrıguez, *Generalized Linear Models* (2017) <http://data.princeton.edu/wws509/r/overdispersion.html>.

included within a publication, can be used or referred to. For example, it is possible that academic software developers are engaging in a more strategic pattern of citation, where they use highly useful patented methods without referring to those articles directly cited by intellectual property rights.[94] In addition, these results do not provide information on other aspects of software reuse, such as the number of downloads associated with a particular open source software package.[95] Thirdly, these results do not provide information about the number of open source bioinformatics programs that are not released as a result of applying for a patent, and the earlier research of Stodden and Reich suggests that a pending patent application is a key factor militating against the release of software under an open source licence.[96]

It is also important to note that this dataset contains a relatively small number of articles, when compared with other studies in this field. For negative binomial regression, a larger number of observations is preferable to ensure that data is not over-dispersed.[97] With respect to other patent-publication studies, Murray and Stern identified 340 articles, whilst Magerman, Van Looy and Debackere identified 584 articles for their study.[98] However, Rai, Allison and Sampat's negative binomial regression model relies on 1010 observations, which is slightly less than the number of observations for this particular study.[99] In addition, the two studies above focus on citation counts over a smaller period of time; for example, Murray and Stern focus on citations from 1997 and 1998 to 2002, leading to 1688 article observations.[100] By contrast, this study considers publication impact over 7 years for some articles in the dataset to up to 20 years for other articles. Accordingly, despite the relative size of this data set compared to other studies, 1134 represents a sufficiently large data set to conduct negative binomial regression on.

A further factor that was not considered in the negative binomial regression analysis was the impact of h-index scores on overall citation rates. The h-index is a metric developed by Jorge Hirsch for measuring the relative quality of researchers; a scholar with an index of *h* has published *h* papers, each of which has been cited in other papers at least *h* times.[101].Therefore,

---

[94] Fiona Murray and Scott Stern, above n39, 673.

[95] Philip M. Davis et al., 'Open Access Publishing, Article Downloads, and Citations: Randomised Controlled Trial' (2008) [337] *BMJ* a568 4.

[96] Isabel Rose Reich and Victoria C. Stodden, above n8, 16.

[97] Joseph M. Hilbe, *Negative Binomial Regression* (Cambridge University Press, 2011) 174.

[98] Fiona Murray and Scott Stern, above n39, 661; Tom Magerman, Bart Van Looy and Koenraad Debackere, above n47, 1705.

[99] Arti K. Rai, John R. Allison and Bhaven N. Sampat, above n9, 1543.

[100] Fiona Murray and Scott Stern, above n39, 663-4.

[101] Jorge. E. Hirsch, 'An Index to Quantify an Individual's Scientific Research Output' (2005) 102(46) *Proceedings of the National Academy of Sciences* 16569–16572 16570.

an attempt was made to identify the change in the h-index over time using the h-index statistics available through the Web of Science. The purpose of adding this factor was to provide an accurate assessment of the impact of h-index scores on overall publication rates for individual articles. Unfortunately, the differences in citation patterns between fields means that the h-index cannot be used to compare the productivity of researchers across disciplines.[102] Accordingly, even though bioinformatics is an interdisciplinary field, it is possible that different authors would receive a greater or fewer number of citations depending on their disciplinary affiliation. It is also possible that the deviance between the five percent critical value and the negative binomial regression value was attributable to this effect. Nevertheless, the differences in citation rates between different disciplines may be a confounding factor. The next Chapter will therefore provide a qualitative assessment to frame and identify this potential confounding factor.

Despite these limitations, from these results the following trends emerge with respect to the patent acquisition in each of the three jurisdictions and the acquisition of an open source licence. The negative binomial regression coefficient was determined as follows. For each of the patent acquisition dummy variables, the p value was not statistically significant. Therefore, the null hypothesis for hypothesis 1 (that is, that patenting has no effect on citation) could not be rejected. In other words, these results would also suggest that patents granted by the USPTO, the EPO and the APO do not have either a negative or positive impact on forward citation rates for the associated journals. The negative binomial regression coefficient for the open source variable was determined as follows. For a one unit change (in other words, as the predictor variable of open source licensing changed to 1), the difference of the log of the expected citation count (response variable) is expected to change by 0.9201, given the other predictor variables in the model are held constant. Accordingly, we can reject the null hypothesis for hypothesis 2 (that is, that open sourcing an algorithm has no effect on citation rates) with a p value of 0.01 (with a 99 percent confidence interval between 0.7640 and 1.0770) in favour of the alternative hypothesis.

These results suggest that open source licensing patented software has a positive effect on forward citations, irrespective of whether the patent has been granted. This result appears to contradict other studies that do not reject the null hypothesis with respect to the relationship between open access and forward citation counts for these articles.[103] In part, this effect may be attributable to the 'open' scientific norms that Section 1.4 of Chapter One described as prevailing within bioinformatics research. These cultural characteristics in turn may have the effect of encouraging researchers to ignore software patents when conducting research. Another

[102] As previously identified by (Lutz Bornmann and Werner Marx, 'The h index as a Research Performance Indicator' (2011) 37(3) *European Science Editing* 77–80 77; Anne-Wil Harzing, Satu Alakangas and David Adams, 'hIa: An Individual Annual h-index to Accommodate Disciplinary and Career Length Differences' (2014) 99(3) *Scientometrics* 811–821 812)

[103] Philip M. Davis et al., above n95, 7.

possibility is that more technically significant algorithms are more likely to be released under an open source licence (as occurred with the BLAST algorithms for genome sequencing, which as discussed in Chapter One was an enormous breakthrough in genome sequencing algorithms and has been cited more than 52,000 times). Finally, self-selection and differences in citation rates by field may account for the variation in this model. However, citation rates do not completely explain the different ways that a bioinformatics algorithm may be used, and may miss other measures which may capture the broader societal benefits from the open source distribution of software. The next chapter provides a qualitative research framework to discuss these issues.

5.4 CONCLUSION

This chapter provides two main quantitative patent landscaping techniques for assessing university and research institute patenting in the bioinformatics space. A qualitative analysis of the patent landscape in the US, the EU, Australia and New Zealand provides support for the notion that subfields of bioinformatics that involve use inspired or applied basic research (such as proteomics or systems biology for drug discovery) are more likely to attract patent protection than subfields of bioinformatics that are more closely associated with pure basic research (such as phylogenetics research). These results suggests that open source and patented bioinformatics software (at least for university and research institute software) do not overlap with one another. Finally, a negative binomial examination of patent publication pairs suggests that patenting a bioinformatics technique does not lead to a decrease in the forward citation of a scientific publication which discloses that technique, but instead that open source licensing is more likely to lead to an increase in forward citation rates for the article.

Chapter 6

QUALITATIVE RESULTS: A GROUNDED THEORY APPROACH TO ANALYSING OPEN SOURCE COMPUTATIONAL BIOLOGY AS KNOWLEDGE COMMONS INSTITUTIONS

6.1 INTRODUCTION

This Chapter qualitatively analyses the effects of copyright and patent protection on the development of computational biology software for the purposes of answering Questions 2 and 3 of this thesis. These questions relate to whether bioinformaticians are relying on patent or copyright protection for their work, and whether these rights are perceived as having a positive or negative impact on research. This stage of analysis completes the process of methodological triangulation of comparing doctrinal, quantitative and qualitative results using a grounded theory approach. As discussed previously in Chapter Four, open source software projects are governed by complex institutional arrangements for determining who can participate in development and who can reuse software. Chapter Five provided some limited empirical support for the argument that releasing bioinformatics software under an open source licence can have a positive effect on reuse and diffusion. This positive effect on forward citations does not appear to be affected when formal intellectual property rights such as patents also apply to that software.

However, this quantitative analysis did not reject the null hypothesis that patents have no impact on forward citations. Accordingly, these findings cannot elucidate the reasons why a particular software package was or was not cited. Further, these findings cannot provide information about the reuse of software packages more broadly. In particular, the decisions made by the developers of open source bioinformatics software about what rules govern their projects and how these rules are implemented may affect the success or failure of open source projects. This success or failure may occur through either ongoing use, development and citation, or failure through lack of use, development or citation. This Chapter therefore describes how grounded theory was used to assess how computational biology developers and researchers perceive the institutional arrangements for the governance of open source bioinformatics projects. Grounded theory is a sociological theoretical approach where primary sources of data are used to inductively generate theory about sociological or socio-technical phenomena.[1] Bioinformatics represents an emergent socio-technical phenomenon in that it involves the merger of computer science and molecular biology. Accordingly, a grounded theory approach was selected for this research project to determine how bioinformaticians and researchers perceive copyright and patent protection for open source bioinformatics projects.

This chapter is split into three parts. Section 6.2 describes the grounded theory methodology used in this study to collect the necessary interview data. This interview data was used to address whether developers and researchers perceive copyright or patent protection as positive or negative for open source bioinformatics projects. In particular,

---

[1] Klaas-Jan Stol, Paul Ralph and Brian Fitzgerald, 'Grounded Theory in Software Engineering Research: A Critical Review and Guidelines' (Paper presented at *Proceedings of the 38th International Conference on Software Engineering* , 2016) 122-3.

Section 6.2 outlines a purposive sampling method which builds on the patent landscaping strategy described in Chapter Five. This purposive sampling method was used to identify a set of interviewees who were located at or were affiliated with academic research institutes in the US, the EU, Australia and New Zealand. Each of these interviewees were involved in successful open source computational biology projects. These interviews were also selected because of their involvement in defining the intellectual property management strategies for those projects. These management strategies include the acquisition of patent rights and the dual licensing of open source software. Section 6.2 then describes how, in addition to an inductive thematic analysis, an adaptation of Elinor Ostrom and Sue Crawford's Institutional Grammar Tool (IGT) matrix was used to categorise different institutional arrangements by complexity and type. These institutional arrangements were divided into rules, shared strategies and norms. Further, these institutional arrangements define the set of permitted and forbidden actions available to each individual participating in the development of open source software. Section 6.3 addresses the results of these semi-structured interviews, dividing each of the results by the institutional rule categories described in Frischmann, Madison and Strandberg's Knowledge Commons framework. Section 6.4 discusses the implications of these results for understanding the role that copyright and patents play in open source bioinformatics research. It also contains a discussion of a rule categorisation approach for identifying perceptions of different forms of intellectual property protection in open source bioinformatics.

6.2 GROUNDED THEORY JUSTIFICATIONS

6.2.1 *Background on Grounded Theory*

Grounded theory refers to a systematic sociological approach developed by Barney Glaser and Anselm Strauss in which rich data (usually but not always drawn from qualitative sources) is used to build new theories about a particular sociological phenomenon.[2] Grounded theory was initially developed for anthropological and sociological research fields, such as nursing research. However, grounded theory has been more frequently used in information systems and common pool resource research to study open source software, data sharing and knowledge commons communities. In particular, many of these communities exhibit rapidly evolving norms of governance due to the emergent technologies they are founded upon.[3] A

---

[2] Kathleen M. Eisenhardt, 'Building Theories from Case Study Research' (1989) 14(4) *Academy of Management Review* 532–550 545; Kathy Charmaz, *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis* (SAGE, 2006) 28-9; Barney G. Glaser and Anselm L. Strauss, *Discovery of Grounded Theory: Strategies for Qualitative Research* (Routledge, 2017) 23.

[3] James Howison, *Alone Together: A Socio-technical Theory of Motivation, Coordination and Collaboration Technologies in Organizing for Free and Open Source Software Development* (PhD Thesis, Syracuse University,

common criticism of qualitative research is that the results flowing from qualitative research lack objectivity. Grounded theory attempts to overcome this criticism by using a continual interplay between data collection and analysis to generate new theory during the research process.[4]

Grounded theory (at least as interpreted by Glaser) itself is critiqued on the grounds that it effectively requires a researcher to start without any theoretical framework when creating new theory.[5] However, Strauss argues that, contrary to Glaser's grounded approach, that a carefully constructed grounded theory approach must have a theoretical perspective to guide the researcher in abstracting significant data categories.[6] An additional key criticism of grounded theory is that it generates highly context sensitive theory that can be difficult to translate to other sociological situations. This context specificity can make that theory difficult to verify.[7] However, Cathy Urquhart and Walter Fernandez argue that grounded theory research can be structured around other research frameworks. This approach to grounded theory does not necessarily require adopting a top down approach where a conceptual scheme is developed and then fieldwork is conducted to confirm its value.[8] In addition, Urquhart and Fernandez argue that grounded theory can be flexibly combined with other research methods and theoretical frameworks to produce more translatable theory than pure grounded theory.[9]

The carefully constructed grounded theory approach used in this thesis has three distinct features. First, this thesis uses theoretical sampling through a preliminary grey literature review. Secondly, the qualitative methodology of this thesis involves making of constant comparisons between different elements of qualitative and quantitative data collected for this thesis. Thirdly, this thesis uses a coding paradigm to understand the sociological phenomenon in question. Specifically, this thesis employs a grounded theory design to compare the institutional intellectual property and technology transfer strategies that are used to govern open source computational biology projects and institutes. Semi-structured interviews are used to collect data about the perspectives of different computational biology researchers. The next section discusses the first stage of this grounded theory approach; that is, theoretical

---

2009) 31; Robert D. Macredie and Kabiru Mijinyawa, 'A Theory-Grounded Framework of Open Source Software Adoption in SMEs' (2011) 20(2) *European Journal of Information Systems* 237–250 245.

4 Glenn A. Bowen, 'Grounded Theory and Sensitizing Concepts' (2006) 5(3) *International Journal of Qualitative Methods* 12–23 13.

5 Cathy Urquhart and Walter Fernández, 'Using Grounded Theory Method in Information Systems: the Researcher as Blank Slate and Other Myths' (2013) 28(3) *Journal of Information Technology* 224–236 226.

6 Barney G. Glaser and Anselm L. Strauss, above n2, 3.

7 Klaas-Jan Stol and Brian Fitzgerald, 'Theory-Oriented Software Engineering' (2015) 101() *Science of Computer Programming* 79–98 91.

8 Cathy Urquhart and Walter Fernández, above n5, 232.

9 Cathy Urquhart and Walter Fernández, above n5, 230-1.

sampling of different open source computational biology communities.

6.2.2 *Theoretical Sampling*

The purpose of theoretical sampling is to ensure that any grounded theory is developed from a representative sample.[10] The first stage of the theoretical sampling strategy used in this thesis was to doctrinally analyse divergent copyright and patent laws, described in Chapters Two and Three of this thesis. The second stage of this theoretical sampling strategy involved comparing the idiosyncrasies in institutional and technology transfer policy, which were described in Chapter Four. In addition, Chapter Four contained a methodological comparison of commons based studies of other socio-technical systems, including open source software and biological data sharing arrangements. In addition to charting the legal and sociological boundaries of socio-technical commons based research, these doctrinal and theoretical systematic reviews were also used to construct the theoretical sampling model adopted in Chapter Five.[11]

A key issue emerging from this initial systematic analysis was the question of how open source software projects and development teams should best be defined as institutions. This definition helps determine the ownership and governance of the common resources under the control of the community. Previous research studying knowledge commons communities has focussed on the dilemma of how to define an institute for the purpose of the IAD framework, and in particular whether formal institutes or communities can be studied. Tania Bubela and colleagues chose to focus on two international mouse genomic data consortiums, the International Knockout Mouse Consortium (IKMC) and the International Phenotypic Mouse Consortium (IPMC). Bubela and colleagues used these institutes to create a grounded theory of mouse genotype and phenotype data sharing.[12] Similar global consortiums have been established for bioinformatics software and data sharing, such as the Genomic Standards Consortium. However, other bioinformatics projects may be significantly more focused and contained within an individual institution.[13] Accordingly, whilst a case study approach was suitable for analysing the IKMC and IPMC, institutionally grounded case studies for bioinformatics would generate a grounded theory that was too context specific.[14] Instead, this

---

[10] Siobhán O'Mahony, 'Guarding the Commons: How Community Managed Software Projects Protect their Work' (2003) 32(7) *Research Policy* 1179–1198 1183.

[11] Mike Weed, '"Meta Interpretation": A Method for the Interpretive Synthesis of Qualitative Research' (2005) 6(1) *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 4-5.

[12] Tania M. Bubela et al., 'Governance of Biomedical Research Commons to Advance Clinical Translation: Lessons from the Mouse Model Community' in Katherine J. Strandburg, Brett M. Frischmann and Michael J. Madison (eds.), *Governing Medical Knowledge Commons* (Cambridge University Press 2017)222 226-27.

[13] Jorge L. Contreras, 'Implementing Procedural Safeguards for the Development of Bioinformatics Interoperability Standards' (2012) 39(2) *Northern Kentucky Law Review* 87–118 110.

[14] Georg von Krogh, Sebastian Spaeth and Karim R Lakhani, 'Community, Joining, and Specialization in Open

thesis attempts to inductively build a grounded theory on open source licensing within bioinformatics.

Because patent rights (and to a lesser degree copyright) are territorial, their effects may vary depending on not only who is applying for a patent but also where the patent is enforced.[15] For example, the Australian innovation system has undergone a transition away from a publicly funded national innovation system headed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) towards a multiplicity of different public research organisations.[16] This model is closer to the institutional variation observed in the US and the EU. As a result, there may be significant regional differences as to the objectives of different researchers, particularly within Australian and New Zealand.[17] Further, within different open source projects, Schweik's research suggests that larger projects are more likely to feature a hierarchical organisational structure governed by the rules surrounding technology transfer.[18] In contrast, projects sponsored by smaller institutes or development groups have been characterised by Clark Asay as being typified by the avoidance of registered intellectual property rights. This avoidance is due to the additional costs associated with acquiring such rights, as well as searching for other competing rights.[19] These smaller institutes may have a lesser sufficient profit incentive or profit-driven organisational culture than their larger compatriots. By contrast, larger institutes might see greater potential in the commercialisation of software, and try to seek patents for their software.[20]

Accordingly, ethics approval was sought from the University of Tasmania Social Sciences Research Ethics Committee (HREC) to conduct up to 50 interviews with bioinformatics researchers in the US, the EU, Australia and New Zealand. The maximum sample size of 50 was chosen as although grounded theory sampling is hypothetically designed to continue until saturation, most grounded theorists recommend conducting between 20 to 50 in-depth grounded theory interviews, with the findings of 25 or more interviews being sufficient to

---

Source Software Innovation: a Case Study' (2003) 32(7) *Research Policy* 1217–1241 1226.

[15] Aexel Metzger, 'Internationalisation of FOSS Contributory Copyright Assignments and Licenses: Jurisdiction-Specific or "Unported"?' (2013) 10(2) *SCRIPTed* 177 190; Ali K. Yetisen and Lisa R. Volpatti, 'Patent Protection and Licensing in Microfluidics' (2014) 14(13) *Lab on a Chip* 2217–2225 2219.

[16] Garrett Upstill and Thomas H. Spurling, 'Adjusting to Changing Times: CSIRO Since the 1970s' (2007) 9(2) *Innovation* 113–124 119.

[17] Mark Dodgson et al., 'Systems Thinking, Market Failure, and the Development of Innovation Policy: The Case of Australia' (2011) 40(9) *Research Policy* 1145–1156 1149.

[18] Charles Schweik and Robert English, 'Preliminary Steps Toward a General Theory of Internet-Based Collective-Action in Digital Information Commons: Findings From a Study of Open Source Software Projects' (2013) 7(2) *International Journal of the Commons* 234–254 239-40.

[19] Clark D. Asay, 'Enabling Patentless Innovation' (2014) 74(3) *Maryland Law Review* 431 434.

[20] Charles M. Schweik and Robert C. English, *Internet Success: A Study of Open-Source Software Commons* (MIT Press, 2012) 43.

challenge existing theories in a particular field.[21]

6.2.3 *Sampling Results*

Out of the potential interviewees identified as appropriate interview targets and contacted by email to participate in this study, 32 interviewees agreed to be interviewed. 30 interviews were finally included (with two being excluded because the interviewees responded via email) in the data that was used in this chapter as these results were sufficient to achieve saturation.[22] Identifying appropriate interviewees itself represented a challenge; the systematic review conducted for this chapter suggested that only a minority of open source software developers will have been faced with potential copyright or patent disputes during the course of their careers.

Accordingly, a list of potential interviewees was drafted using a purposive sampling approach. Firstly, potential interviewees were identified using both granted patents and ongoing patent applications out of the list of patent inventors. This list included inventors who had prior to or after filing their patents released the patented algorithm under an open source software licence.[23] This list was also curated to include bioinformaticians and computer scientists who worked at the institutes and had coauthored papers with patent applicants. This decision was justified on the grounds the patent applicants themselves may not be the only parties responsible for developing patented open source projects.[24] Though small in size, this list was useful in helping to identify interviewees who could help elucidate the conflict over deciding between these two forms of software release.[25] Efforts were made to increase social

---

[21] Kathy Charmaz, above n2, 144; John W. Creswell, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (SAGE Publications, 2012) 86; Bryan Marshall et al., 'Does Sample Size Matter in Qualitative Research?: A Review of Qualitative Interviews in IS Research' (2013) 54(1) *Journal of Computer Information Systems* 11–22 13.

[22] There is no explicit limit on the maximum number of interviews that should be conducted for qualitative research. Mark Mason suggests that a minimum of 8 interviews may be sufficient to draw conclusions about a sample population and reach saturation. However, Mason notes that for qualitative studies, the average number of interviews is 31, with non normal data distribution; in other words, the majority of semi structured interviewed studies involve 30, 40 or 50 interviewees (Mark Mason, 'Sample Size and Saturation in PhD Studies Using Qualitative Interviews' (2010) 11(3) *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 6-7)

[23] Dietmar Harhoff, Frederic M Scherer and Katrin Vopel, 'Citations, Family Size, Opposition and the Value of Patent Rights' (2003) 32(8) *Research Policy* 1343–1363 1349.

[24] Nicole Ziegler, Oliver Gassmann and Sascha Friesike, 'Why do Firms Give Away their Patents for Free?' (2014) 37(2) *World Patent Information* 19–25 3; Robert P. Merges and Michael Mattioli, 'Measuring the Costs and Benefits of Patent Pools' (2017) 78(2) *Ohio State Law Journal* 281–348 307.

[25] Kevin J. Boudreau and Karim R. Lakhani, "Open" Disclosure of Innovations, Incentives and Follow-On Reuse: Theory on Processes of Cumulative Innovation and a Field Experiment in Computational Biology' (2015) 44(1) *Research Policy* 4–19 5.

desirability of participation by offering to de-identify all interview transcripts.[26] However, after a draft initial list was compiled the author concluded that patent applicants who had released open source software might be unwilling to conduct interviews. This unwillingness could be attributable to ongoing patent applications or commercial licensing of software.[27] This effect may be particularly true given that Victoria Stodden reports that computer science researchers are particularly inclined to be biased towards Mertonian norms, and filing for patents might be seen as contrary to those norms.[28] Accordingly, this initial list of interviewees who had filed for bioinformatics patents was expanded to include additional researchers associated with key open source projects in bioinformatics research.

Interviewee names and email addresses were collected via searching for articles on Scopus, as well as a scoping review to see whether they had publicly commented on the role of patents and copyright protection in open scientific research or open source licensing. This sampling strategy captured any researchers who had had negative experiences with respect to patent or copyright enforcement for bioinformatics software. In particular, this strategy was designed to capture information that had not been documented in formal legal literature but may nevertheless impact on motivations to engage in an open source project.[29] In addition, sampling in this fashion helped develop a wide list of open source projects and developers, which in turn helped demonstrate how *in vitro* and *in silico* researchers interacted with one another in an open innovation context. Twenty six of these interviews were conducted using a Voice over IP (VoIP) client (Skype, Google Hangouts or Tox).[30] Four of these interviews were conducted via a telephone call. Two interviews were conducted face to face. Interview invitations were sent in 2017. Each interview was conducted by the author, and ranged from 35 minutes to 80 minutes, with an average time of 42 minutes. Although the author prepared an interview schedule, interviewees were given the opportunity to set the frame of the discussion in accordance with a semi-structured interview approach. Interviewees were also allowed to explain concepts that would not otherwise have been available with a rigid question and answer approach.[31] Non-respondents were sent a follow up email two weeks after the

---

[26] Anja Schoen, Bruno van Pottelsberghe de la Potterie and Joachim Henkel, 'Governance Typology of Universities' Technology Transfer Processes' (2014) 39(3) *The Journal of Technology Transfer* 435–453 445.

[27] Arti K. Rai, John R. Allison and Bhaven N. Sampat, 'University Software Ownership and Litigation: A First Examination' (2009) 87(5) *North Carolina Law Review* 1519–1570 1546.

[28] Victoria Stodden, 'The Scientific Method in Practice: Reproducibility in The Computational Sciences' (MIT Sloan School Working Paper No 4773-10, MIT Sloan School of Management, 2010) 26 <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1550193>.

[29] Marcus M. Dapp, *The Effects of Software Patent Policy on the Motivation and Innovation of Free and Open Source Software Developers* (PhD Thesis, ETH Zurich, 2009) 55-6.

[30] Hannah Deakin and Kelly Wakefield, 'Skype Interviewing: Reflections of Two PhD Researchers' (2014) 14(5) *Qualitative Research* 603–616 606.

[31] 'Empirical Research in International Management: A Critique and Future Agenda' in Rebecca Marschan-Piekkari

initial email to ask for their participation. After each interview, the recordings of each were transcribed by the author and sent back to the interviewee to confirm that they were satisfied with the contents and that any confidential information had been redacted.

Interviewees working in certain fields of research were reluctant to participate in this study. As the systematic reviews revealed, a number of companies (including Microsoft, Illumina and International Business Machines (IBM)) have significant ties to certain research institutes and universities.[32] These public-private research connections meant that patentees who were identified using the patent landscaping strategy described above were often reluctant to be interviewed due to ongoing patent applications or work commitments. Once these early rejections were received, the author ceased sending invitations to researchers working on private sector or on joint public-private sector research projects. In addition, researchers at some universities with a high number of software patents either refused to be interviewed or did not reply to requests to be interviewed. As a result, it was not possible to understand how these institutes managed the nexus between patenting and open source licensing.[33]

As articulated in the comparative section of Chapter Five, the first stage of constant comparison using a grounded theory approach involved comparing each of the interviewees by jurisdiction as described in table 6.1.

Jurisdiction	**Number of Interviewees**	**Percentage of Interviewees**
US	12	40.00%
EU	10	33.34%
Australia and New Zealand	8	26.67%

Table 6.1: Interviewee numbers by jurisdiction

Interviewees were also classified by their disciplinary background, as set out in Table 6.2:[34]

These researchers were then interviewed using a semi-structured interview schema, which provided scope for the interviewer to ask unique questions about the interviewee's engagement

---

and Catherine Welch (eds.), *Handbook of Qualitative Research Methods for International Business* (Edward Elgar 2004)25 34-5.

[32] *University of Washington Department of Genome Sciences Joins Illumina Genome Network* (1 July 2011) TechnologyNetworks <https://www.technologynetworks.com/genomics/news/university-of-washington-department-of-genome-sciences-joins-illumina-genome-network-213020>

[33] Arti K. Rai, John R. Allison and Bhaven N. Sampat, above n27, 1547-8.

[34] Andrew Bartlett, Jamie Lewis and Matthew L. Williams, 'Generations of Interdisciplinarity in Bioinformatics' (2016) 35(2) *New Genetics and Society* 1–24 8.

Academic Background	Number of Interviewees	Percentage of Interviewees
Computer Science and Bioinformatics	12	37.50 %
Statistics and Mathematics	6	18.75%
Molecular Biology	9	31.24%
Organic Chemistry	2	9.38%
Physics	1	3.13%

Table 6.2: Interviewee breakdown by academic background (that is, the discipline of the highest degree which each interviewee received).

with the patent system.[35]

6.2.4 *Comparative Approaches to Patenting, Open Source Licensing and Technology Transfer*

Jamie Lewis and Andrew Bartlett describe how there are interdisciplinary tensions in bioinformatics research. Specifically, these institutional tensions exist between researchers from molecular biology backgrounds against those from computer science or mathematics backgrounds.[36] Lewis and Bartlett's qualitative interviews with bioinformaticians in the UK explored how molecular biologists often view bioinformatics software as a service dependent on primary data produced through *in vitro* research. By contrast, bioinformaticians view bioinformatics as a scientific discipline it its own right.[37] Lewis and Bartlett's research was conducted in the context of examining the evolution of bioinformatics as a field. However, their subsequent study with Matthew Williams revealed that those who viewed patents as an indicator of esteem in research were statistically more likely to see bioinformatics as a service.[38] The interview transcripts in the present study were coded to compare how researchers of different institutional backgrounds perceived the utility patents technology transfer offices. In particular, it was expected that molecular biology researchers would have a stronger perception of technology transfer than those researchers with a background in computer science or mathematics.[39] This approach to comparative coding was underpinned by earlier qualitative research conducted by Andrew Harvey and Mark McMeekin into the

---

[35] Harrie Jansen, 'The Logic of Qualitative Survey Research and its Position in the Field of Social Research Methods' (2010) 11(2) *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*.

[36] Jamie Lewis and Andrew Bartlett, 'Inscribing a Discipline: Tensions in the Field of Bioinformatics' (2013) 32(3) *New Genetics and Society* 243–263 250.

[37] Jamie Lewis and Andrew Bartlett, above n36, 256.

[38] Andrew Bartlett, Jamie Lewis and Matthew L. Williams, above n34, 13.

[39] Wan-Ling Huang, Mary K. Feeney and Eric W. Welch, 'Organizational and Individual Determinants of Patent Production of Academic Scientists and Engineers in the United States' (2011) 38(6) *Science and Public Policy* 463–479 475.

development of proprietary bioinformatics software such as Lion BioSciences. Their findings into the strong 'open source' culture in bioinformatics and opposition to patent protection are consistent with the findings presented in Chapter Five.[40]

An additional level of comparison in the present study involved considering the scale of commons regimes and the effects of divergence in national copyright laws, patent regimes and policies on technology transfer.[41] These high level rules amounted to 'constitutional rules' on how intellectual property sharing regimes should be best implemented by universities and research institutes.[42] Moreover, the doctrinal analysis from Chapters Two and Three established the high level differences in copyright and patent regimes between the US, the EU, Australia and New Zealand. This doctrinal analysis demonstrated that with respect to patent protection, the US has historically adopted a more liberal approach to software and diagnostic method patenting than the EU, Australia and New Zealand. On the other hand, the fair dealing or private use exceptions in EU, Australian and New Zealand copyright law have been more narrowly interpreted than the fair use norm in US copyright law.[43]

6.2.5 *Coding Metric - Using the Institutional Grammar Tool (IGT)*

The next stage of the grounded theory approach applied in this thesis involved using a coding mechanism to compare the responses that different interviewees gave on their perspectives towards copyright and patent protection in bioinformatics. These raw results were then coded using RQDA, a qualitative data analysis library for the R statistical computing platform.[44] In addition to thematic coding,[45] the results were also coded using the Institutional Grammar Tool (IGT). The IGT was first developed by Ostrom and Sue Crawford to enable empirical analysis

---

[40] Mark Harvey and Andrew McMeekin, 'Public or Private Economies of Knowledge: The Economics of Diffusion and Appropriation of Bioinformatics Tools' (2009) 4(1) *International Journal of the Commons* 481–506 495.

[41] David W. Cash et al., 'Scale and Cross-Scale Dynamics: Governance and Information in a Multilevel World' (2006) 11(2) *Ecology and Society* 2; Eleanor Flening, *30 Years After the Bayh-Dole Act: Rethinking the Australian Research Commercialisation Experience* (PhD Thesis, 2012) 178-9 <https://digitalcollections.anu.edu.au/handle/1885/9055>; Jerome H. Reichman, Tom Dedeurwaerdere and Paul F. Uhlir, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge University Press, 2016) 325-6.

[42] Charlotte Hess and Elinor Ostrom, 'A Framework for Analyzing the Knowledge Commons' in Charlotte Hess and Elinor Ostrom (eds.), *Understanding Knowledge as a Commons: from Theory to Practice* (MIT Press 2007) 50.

[43] Jerome H. Reichman and Ruth L. Okediji, 'When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale' (2012) 96(4) *Minnesota Law Review* 1362–1480 1434; Alexandra Sims, 'The Case for Fair Use in New Zealand' (2016) 24(2) *International Journal of Law and Information Technology* 176–202 179; Ezieddin Elmahjub and Nicolas P. Suzor, 'Fair Use and Fairness in Copyright: A Distributive Justice Perspective on Users' Rights' (2017) 43(1) *Monash University Law Review* 274–298 276.

[44] Ronggui Huang, *RQDA: R-based qualitative data analysis* (2018) <http://rqda.r-forge.r-project.org/>.

[45] Virginia Braun and Victoria Clarke, 'Using thematic analysis in psychology' (2006) 3(2) *Qualitative Research in Psychology* 77–101.

of how rules control and affect institutional behaviour.[46] However, Geary exclusively focuses on the use of institutional logics in the governance of an international commons based resource (namely the DNA barcode commons, which is a consortium of different scientific research institutes developing a taxonomic coding system for biodiversity). The IGT in this study is used to examine institutional compliance in the context of technology transfer across a number of different institutional settings that produce bioinformatics software. Using Ostrom's in use classification structure for governance arrangements described previously[47], the IGT divides institutional statements for governance into three categories;

1. rules;
2. norms; and
3. shared strategies.

Using the IGT model, institutional arrangements are classified into these three categories based on the presence of five grammatical components; *'Attributes'*, *'Deontics'*, *'Aims'*, *Conditions* and *'Or else'* statements, forming the acronym ADICO.[48] These grammatical components, described in Table 6.3, can be combined to form the three categories of institutional statement described above. Firstly, shared strategies can be written as *'Attributes'*, *'Aim'* and *'Conditions'* (AIC). Secondly, norms can be written as *Attributes*, *Deontic*, *Aim* and *Conditions* (ADIC). Thirdly, all rules can be written as *Attributes*, *Deontic*, *Aim*, *Conditions* and *Or Else* (ADICO).

Crawford and Ostrom initially applied the IGT to policy statements, although Xavier Basurto and colleagues have expanded it to be used to analyse the grammatical structure of legislative statements (using the example of US transportation policy).[49] Basurto and colleagues included an extra element to the deontic component of the IGT, recognising that deontic statements could be implicit as well as explicit.[50] In addition, Basurto, Siddiki and Weible compared the coding of institutional statements with thematic coding of

---

[46] Sue E. S. Crawford and Elinor Ostrom, 'A Grammar of Institutions' (1995) 89(3) *American Political Science Review* 582–600.

[47] Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, 1990) 38.

[48] Sue E. S. Crawford and Elinor Ostrom, above n46, 538.

[49] Xavier Basurto et al., 'A Systematic Approach to Institutional Analysis: Applying Crawford and Ostrom's Grammar' (2010) 63(3) *Political Research Quarterly* 523–537 524.

[50] Elinor Ostrom and Sue E. S. Crawford, 'Classifying Rules' in Elinor Ostrom (ed.), *Understanding Institutional Diversity* (Princeton University Press 2009)168 171; Xavier Basurto et al., above n49, 525.

Attributes	*Attributes* is a holder for any value of a participant-level variable that distinguished to whom the institutional statement applies. For an individual the attribute may be the individual's position within their organisation. For an organisation, the attribute may be size or the organisation's relationship to other organisations.
Deontic	*Deontic* is a holder for the three modal verbs using deontic logic: *may* (permitted), *must* (obliged), and *must not* (forbidden). The deontic applies to the attributes of an individual or organisation.
A**i**m	*Aim* is a holder that describes particular actions or outcomes to which the deontic is assigned.
Conditions	*Conditions* is a holder for those variables which define when, where, how, and to what extent an *Aim* is permitted, obligatory, or forbidden.
Or Else	*Or Else* is a holder for those variables which define the sanctions to be imposed for not following the rule.

Table 6.3: Sue E. S. Crawford and Elinor Ostrom, 'A Grammar of Institutions' (1995) 89(3) *American Political Science Review* 582–600.

semi-structured interviews. Specifically, institutional members were asked whether they would or would not comply with each institutional statement to draw a link between rules in form and implicit rules in use.[51] For this study, a similar focus was drawn to compliance with institutional rules to determine if interviewees were bypassing the technology transfer office at their institutions.[52]

An additional step in the IGT involves matching different institutional statements to the elements of a commons that they govern (otherwise known as horizontal coding).[53] To generate institutional categorisations, Geary used the institutional logic theory to categorise institutional statements generated through semi-structured interviewing. This coding involved identifying community policies and exogenous factors by their type and nature (specifically corporation, market, profession and state).[54] Likewise, this thesis relies on semi-structured interview analysis (as opposed to institutional document analysis) as a measure of assessing the rules and exogenous factors that are required to govern open source bioinformatics

[51] Saba Siddiki, Xavier Basurto and Christopher M. Weible, 'Using the Institutional Grammar Tool to Understand Regulatory Compliance: The Case of Colorado Aquaculture' (2012) 6(2) *Regulation & Governance* 167–188 170-1.

[52] Rajeev K. Goel and Devrim Göktepe-Hultén, 'What Drives Academic Patentees to Bypass TTOs? Evidence From a Large Public Research Organisation' (2017) *The Journal of Technology Transfer* 1–19 12-3.

[53] Elinor Ostrom and Sue E. S. Crawford, above n50, 187.

[54] Janis Dawn Geary, *How the DNA Barcode Commons are Governed: Understanding how a Heterogeneous Global Community Shares Genetic Resources for Non-Commercial Use* (PhD Thesis, University of Alberta, 2017) 82, 95.

research. However, Geary combines the institutional grammar tool with institutional logics to analyse the effect of heterogeneity in the formation of a knowledge commons. For broader comparative studies, Ostrom and Crawford recommend explicitly linking the IAD rule types (which were discussed in Section 4.3.4 of Chapter Four) and the IGT institutional statements. This recommendation is based on the connection between institutional grammar and the general IAD framework.[55] Accordingly, this thesis combines the IGT rule types with the IAD rule types to understand the role of different rules in the management of a bioinformatics project. These different rule types are explained in further detail in Table 6.4.

[55] Elinor Ostrom and Sue E. S. Crawford, 'A Grammar of Institutions' in Elinor Ostrom (ed.), *Understanding Institutional Diversity* (Princeton University Press 2009)137–174 138; Elinor Ostrom and Sue E. S. Crawford, above n50, 187-8; Aaron M. Lien, Edella Schlager and Ashly Lona, 'Using institutional grammar to improve understanding of the form and function of payment for ecosystem services programs' (2018) 31() *Ecosystem Services* 21–31 22-3.

Position Rules:	Define the actors in an action situation, as well as the number of actors (quorum) required to make a particular decision. The generic aIm verb for position rules is 'be'.
Boundary Rules:	Set the conditions under which an actor may occupy a given position. Boundary rules may either be credential based or procedure based. The generic aIm verbs for boundary rules relates to entry or exit.
Choice Rules:	Specify what a participant must, may or must not do at a particular point within the decision process. The aIm verb relate to whether this action is mandatory, optimal or forbidden.
Information Rules:	Define how information is exchanged between actors in an institution. Information rules focus on how information is shared, what information is shared and when information sharing takes place. The generic aIm verbs are send or receive.
Payoff Rules:	Establish rewards and sanctions for actors under certain conditions. These statements determine how much an ecosystem service is worth but also set penalties for failing to maintain the shared resource. The generic aIm verbs are pay or receive.
Aggregation Rules:	Define how decisions are made between actors. These rules can address any decision process that involves more than one actor. The generic aIm verb will vary but usually relate to rules that require the joint approval of one or more actors.
Scope Rules:	All other rules that cannot be classified into one of the rule types described above will fall into scope rule types, These rules dictate the actions that participants may or may not take, and the outcomes that flow from those decisions. The general aIm for scope rules is 'occur'.

Table 6.4: Ostrom and Crawford's rule categories, along with an analogy to these rule categories for a default open source project. Adapted from Sue E. S. Crawford and Elinor Ostrom, 'Classifying Rules' in Elinor Ostrom (ed), *Understanding Institutional Diversity* (Princeton University Press, 2005) 186, 191 and Charles Schweik and Robert English, *Internet Success: A Study of Open Source Software Commons* (MIT Press, 2012), 92

The different rule types and institutional statements were first identified through deductive coding, which involves grouping rules by type and by institutional statement categories. Inductive coding then involves developing a set of specific sub rule types or themes within the existing IAD rule types from the data.[56]

6.3 QUALITATIVE RESULTS

6.3.1 *Institutional Statement Classifications*

6.3.1.1 Position Rules

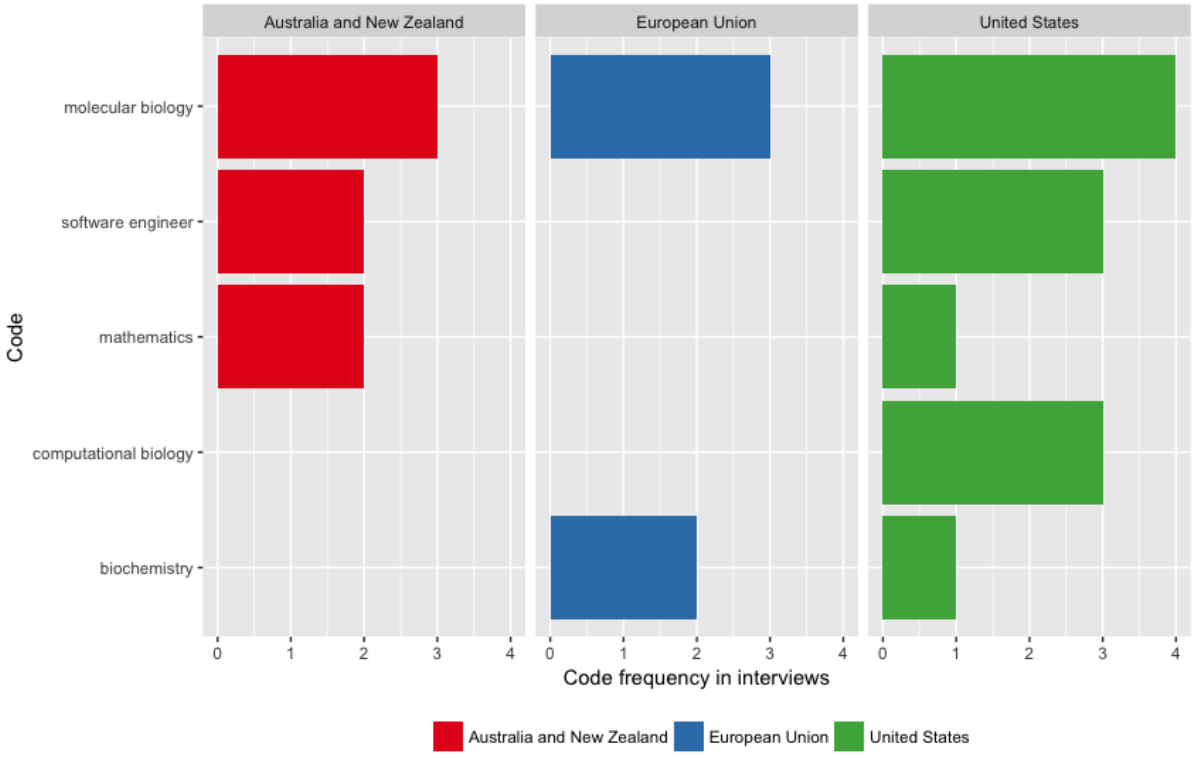

Figure 6.1: Position rules identified through interview transcription.

From the position rules identified through inductive coding, three positions were identified with respect to the different roles that researchers can occupy within an open source project. As James Howison and James Herbsleb have discussed, the emergence of new forms of collaborative tools has created the formation of new organisational subgroups in

[56] Ann R. R. Robertson et al., 'Tightrope walking towards maximising secondary uses of digitised health data: a qualitative study.' (2016) 23(3) *Journal of Innovation in Health Informatics* 591–599 593; R. Stuart Geiger et al., 'The Types, Roles, and Practices of Documentation in Data Analytics Open Source Software Libraries' (2018) 27(3) *Computer Supported Cooperative Work (CSCW)* 767–802 770.

bioinformatics development.[57] Specifically, the following three groups were identified in the present study (although it should be noted that interviewees often held overlapping responsibilities depending on the size of the project and goals of the project):

1. Molecular biologists, zoologists and medical scientists who collect primary sequence data using *in vitro* sequencing techniques
2. Statisticians and mathematicians, who are responsible for cleaning any primary sequence data and preparing it for analysis
3. Bioinformatics software developers, who write software to analyse primary sequence data and receive data from data scientists

These categories identified in the present study also correlate to Jamie Lewis and Andrew Bartlett's distinction between primary and secondary inscriptions of data and the researchers who are responsible for each category.[58] Each of these groups may have competing considerations surrounding the role of intellectual property rights. Some researchers viewed intellectual property rights not only as a tool for commercialisation of research results but also as a means of ensuring the recognition of work. These competing considerations were revealed in the interviews, with one interviewee in particular describing their frustration at the disciplinary divides within bioinformatics research in the quote below:

> *My perspective is that genomic medicine is bringing challenges to computational biology which are not present in conventional statistics... We are measuring a massive number of parameters on a massive number of subjects. There is a real bottleneck at the moment in the genomic revolution and this bottleneck is largely to do with algorithm performance'. However, if you present these problems to a biologist, then they will try and focus on their immediate research task.* - Australian Researcher

Nevertheless, there was still substantial overlap between each positions, with the vast majority of interviewees involved in both setting original research goals and developing software to support other research goals. This finding was particularly true for smaller open

---

[57] James Howison et al., 'Understanding the Scientific Software Ecosystem and its Impact: Current and Future Measures' (2015) 24(4) *Research Evaluation* 454–470 455-8.

[58] Jamie Lewis and Andrew Bartlett, above n36, 245; Andrew Bartlett et al., 'The Locus of Legitimate Interpretation in Big Data Sciences: Lessons for Computational Social Science rom -Omic Biology and High-Energy Physics' (2018) 5(1) *Big Data & Society* 2053951718768831 4.

source projects, which were developed as part of individual research projects. By contrast, larger open source projects were characterised by the involvement of more specialised bioinformatics developers being involved with the project. The overlap between each of these competing considerations is discussed below when considering the boundary rules associated with each project.

6.3.1.2 Boundary Rules

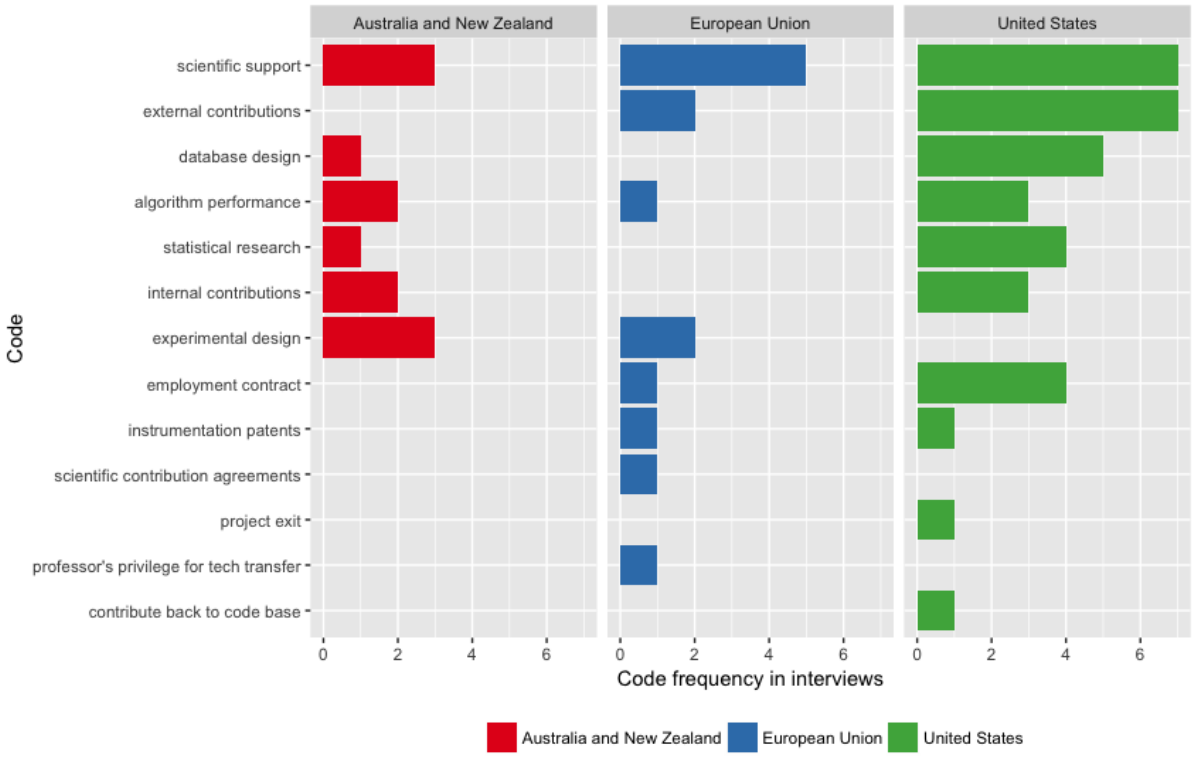

Figure 6.2: Boundary rules identified through interview transcription.

Following inductive coding, the vast majority of boundary rules related to the sharing of data for the production of bioinformatics software. As alluded to in Section 1.5 of Chapter One, a historical account of the evolution of bioinformatics reveals a conflict between producing reproducible software versus ensuring the sustainability of scientific software development. Four US interviewees and one European interviewee noted that one of the early strategies adopted by government funded genomics researchers was rapid release of data. This strategy was designed to create a prior art base and thereby prevent private or public-private consortiums from acquiring patent protection.[59] However, the introduction of new software engineering tools have radically changed the development of bioinformatics software. For

[59] Hallam Stevens, 'The Politics of Sequence: Data Sharing and the Open Source Software Movement' (2015) 50(4) *Information & Culture: A Journal of History* 465–503 467.

example, the introduction of content versioning tools such as GitHub, as discussed in Section 1.4.3 of Chapter One, have significantly enhanced the ability of developers to collaborate on a distributed software project.[60] This strategy has heavily influenced how grant agencies allocate funding for computational biology research. In turn, these strategies have influenced the perspectives of bioinformatics researchers, particularly with respect to the notion that commercialising publicly funded research innovation amounts to 'double taxation':[61]

> *In my group we try to make everything as open as humanly possible. Because I know we are being paid by the taxpayer so I believe the legislation which requires academics to take responsibility and for me that means we need to do the best for the population at large. Even beyond New Zealand. Science is meant to be a common good. If I'm getting paid by taxpayers to do my research at minimum the taxpayer should be able to see what I have been doing* - New Zealand Researcher.

However, several researchers across jurisdictions mentioned that a pure government funding model directly conflicted with sustainability, for both data and software:

> *(The whole field of bioinformatics) concerns itself with data transformation... without being derogatory, there's a lot of hacking of different files together [to view] different associations [between different groups of biological data]* - US Researcher

Although reproducibility of research results is consistently encouraged in bioinformatics development, these statements indicate that open source development is frequently focussed on a particular research project. Once the funding for that project elapses, so too does the funding to support ongoing software development. The cessation of funding then increases the incentive to develop proprietary bioinformatics software. Nevertheless, developers still found ways to ensure that at least part of the software remained available under an open source license:

> *We treat open source software as almost a trial version of what we can offer. If a researcher comes to us with a specific problem that they want to solve with our software, we fork it and develop a bespoke version that they pay for. With the bespoke version, we then attach headers to the bespoke forked version so that if we see this version online, we know that (infringement has occurred)* - New Zealand Researcher

---

[60] Karthik Ram, 'Git can facilitate greater reproducibility and increased transparency in science' (2013) 8(1) *Source Code for Biology and Medicine* 7.

[61] Double taxation refers to the concept that taxpayers already pay once for public research and therefore should not have to do so again (Shubha Ghosh, 'Are Universities Special' (2016) 49(3) *Akron Law Review* 671–694 684).

6.3.1.3 *Choice Rules*

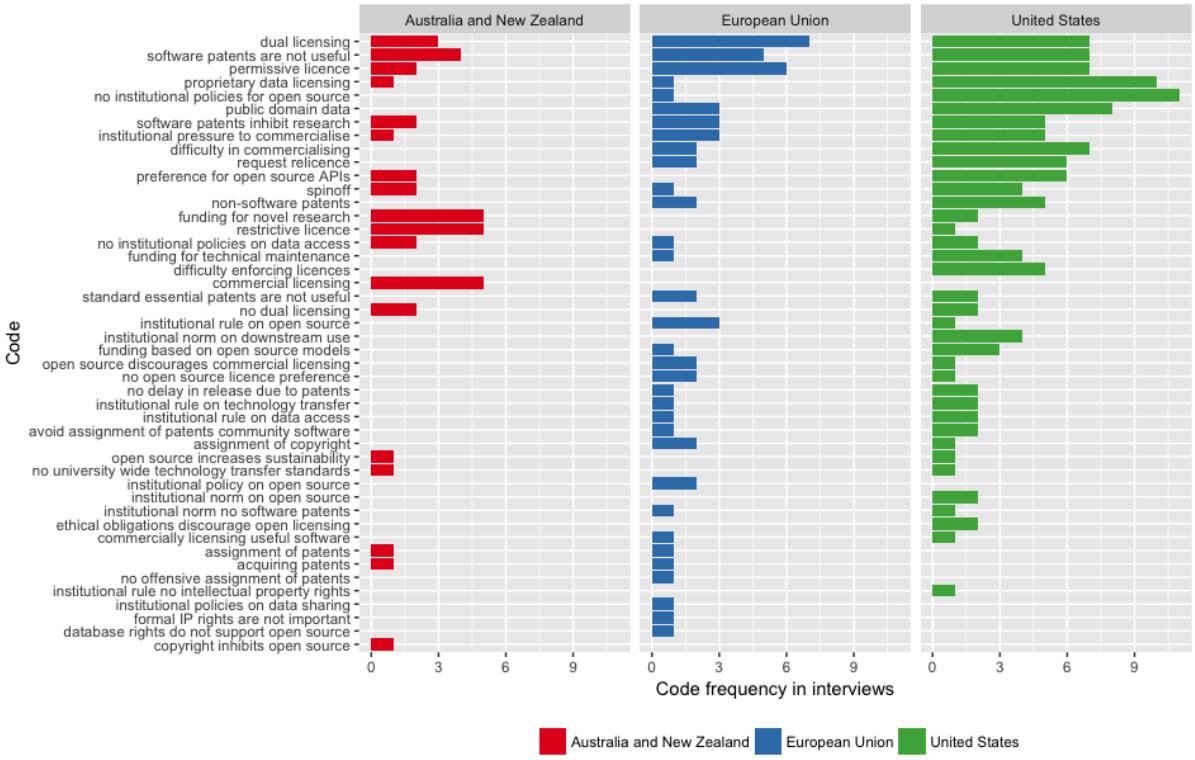

Figure 6.3: Choice rules identified through interview transcription.

The choice of whether to distribute the software was a boundary rule identified by many interviewees (which was a factor identified by one interviewee as resting with the developer). The other choice rules identified in the present study through inductive coding were institutional arrangements relating to open source licensing, patenting and commercialising bioinformatics software. A breakdown of the different licences preferred by interviewees from each jurisdiction is shown below in Table 6.6:

Licence Type	Permissive	Restrictive	Bespoke or Dual	No Strong Preference
Jurisdiction				
US	6	2	0	4
EU	5	4	0	1
Australia and New Zealand	4	3	1	0

Table 6.5: Preference for open source software licences by jurisdiction.

This table demonstrates that whilst there were some interviewees who did not express a preference on open source licences (particularly in the US) a plurality expressed their

preference as permissive licences. Surprisingly few interviewees expressed their preference for restrictive licences, and were prepared to re-licence their project should the opportunity for future commercialisation emerge:

> *Interviewer: What impact do you think that will have on your translation plans, as the GPL version 3 has a provision that prevents downstream licensing?*
>
> *Interviewee: So the version we have now is very much an outgrowth of my initial architecture. Version 2.0 will provide much more general functionality and much more versatility. What we might release that under will depend on the people we have engaged to commercialise the software. So my perspective is that what we have done so far is out in the public domain, but that might not be the case if there is some way of sustaining the software through commercialisation* - Australian researcher

This perspective was coupled with the notion that it is better for software to be published than to languish on a private server and remain undeveloped:

> *I guess the challenge here is to ensure that the software is fit for purpose whilst encouraging developers at the same time to release their software. I have had discussions with a student of mine recently who produced this great software and I asked him whether he had it backed up – he didn't have it backed up, he didn't have it uploaded onto Github. The other side of this is to make sure that source code is fit for purpose. I still have the software from my dissertation at home on my computer. This is the challenge, because whilst a lot of this software performs the function it is required to perform, it is not necessarily the most elegant and efficient code. So those are the key challenges.* - US Research and Developer

In addition to individual preferences on open source licensing, only one European interviewee identified an IGT rule on the choice of open source licences for software published by the institute:

> *Because of the way that [our institute] works, all software [we release] has to pass through a licensing committee, and we have 3 different types of licences that we recommend for staff. For web based services, we recommend Apache licences, so for [project] we use Apache. The GPL is also preapproved but for research code only. The third is a free for academic use/commercial use paid licence, which is promulgated by our technology transfer licence. If someone wanted to use a different licence, they would need to ask for approval and have a good justification.*

With respect to choice rules around patenting, the rationalisations for and against patenting are set out in Table 6.7 above.

	US	**European Union**	**Australia**
Reasons For Patenting	Seeking private investment Protecting Biomedical Research	Protecting Novel and Industrially Useful Research	Encourages Investment in Biomedical Research Provides Funding for Maintenance
Reasons Against Patenting	Difficulty of Enforcement Discourages Reuse Loss of Reputation Cost of Patent Acquisition Difficulty In Applying A Business Model	Cost of Patent Acquisition Loss of First Mover Advantage Inhibits Reuse Discourages Citation Difficulty in Identifying a Business Model	Difficulty in Identifying a Business Model Damage to Reputation Discourages Citation Discourages Reuse

Table 6.6: Reasons for and against patenting broken down by jurisdiction

Only two interviewees from Australia and New Zealand (both computer scientists), two interviewees from the European Union (one computer scientist and one molecular biologist) and three interviewees from the US (all molecular biologists) mentioned that they had applied for or would acquire patents during the course of software development. These findings suggest that propensity to acquire patents in bioinformatics is not influenced by disciplinary background. Further, these interviewees strongly implied that commercialisation only became a consideration once a potentially valuable technique had been identified. This perspective was somewhat reflected in the patent landscaping model described in Chapter Five, where at least one interviewee was significantly less concerned about patents that were directed towards laboratory patents as opposed to software patents:

> *(It's important to) differentiate 'knowledge' patents and 'laboratory' patents... I pick up the opinion that some laboratory patents are very general and have wide applicability and that there are so many 'micropatents' that they normally have to come up for a patent for animal models, a patent for human models, a extraction patent, all the rest of it. If I go to software patents, then I am very strongly against software patents. Software by its very nature builds on other software. The stuff I*

write depends on Apache software and other software... the fact that you build on other work is critical in software and there is a huge [amount] of communal work [that depends on free reuse of components] - European researcher

The translational or clinical applicability of a particular software package was considered an important factor in whether to seek patent protection due to the need for clinical viability of research software. This finding confirmed the patent landscaping model described in Section 5.3.1 of Chapter Five. In particular, this section discussed how use-inspired or applied bioinformatics applications (for drug discovery or personalised medicine) featured in the majority of patent applications and grants. By contrast, basic bioinformatics inventions (such as phylogenetics research software) featured in a significantly smaller number of the patents described in Chapter Five:

(The software projects that are frequently commercialised include those) that are focused on bringing together clinical data, family health data, and 'omics data'; that is, the translational medicine aspect of bioinformatics. All of this is part of a push towards translational and personalised medicine (guided by bioinformatics) - US Researcher

Another key factor that discouraged the acquisition of patents was the large amount of prior art from open source software:

[T]here's a lot of open source software that's available on GitHub, and that software is developed by a graduate student who is doing cutting edge research on genomics, and they do this part of project. Because it's open source, there's no barrier to download the software, regardless of the quality of the software or whether it is fit for purpose. So that was one of the key competitors - the huge amount of free software that was out there to try. - New Zealand Researcher

Interestingly though, for five Australian and New Zealand interviewees, eight European interviewees and eight US interviewees, employers knowingly permitted individual laboratories to determine how they approached the patenting of bioinformatics. Further, some researchers admitted writing a requirement for open source licensing into grant applications. In part, researchers noted that universities and research institutes permitted submission of grants in this form, as patent rights on software did not constitute a major source of revenue for these institutions. This revenue stream can be compared to industry-academic collaborations on matters such as pharmaceutical development. In these situations, the heterogeneity of the organisation could in fact be beneficial for promoting open source licensing:

> *Basically – the way it works [at our institute] is that we use the royalties from patents and biologics that our basic research supports. Then because that money goes back into a common pool, we can use it to fund bioinformatics research. But that only works because of the interdisciplinary nature of (our institute). At the same time though, that is indicative of the benefits of the open source model for bioinformatics. Because in academia our main funding from the government comes through publications, we need to keep software open to publish it. If we publish we get funding, so that supports our ongoing development. And by this, I don't only mean funding the development of new software, I also mean supporting new software* - Australian researcher.

Another important factor that interviewees referred to was resolving institutional pressure to commercialise software with community expectations that bioinformatics software should be openly licensed. On the one hand, working with a technology transfer office represents a strategy by which scientific software developers can guarantee the sustainability of their work. On the other hand, the reverse response can occur when the broader bioinformatics community rebels against the idea of using proprietary software, a phenomenon that has been reported in other open source communities.[62] These observations demonstrate how technology transfer offices, open source software projects and private collaborators may act as competing institutions within a broader organisational framework.[63] Consistent with the doctrinal analysis on legislative mechanisms to encourage technology transfer, the greatest number of researchers who experienced institutional pressure not to release software projects under an open source licence were from the US. In particular, six interviewees reported either technology transfer staff or other researchers encouraging them to patent software. European, Australian and New Zealand interviewees stated that they had not encountered these same pressures at an institutional level. However, one EU interviewee noted that divergences in approaches to technology transfer contradicted official EU policy on openness:

> *It varies on a national level. The UK has generally adopted the 'American' perspective, but it really depends on the country. France has historically been very protecting of its informatics research, but I am not so sure about other countries. That's changing though as the European Commission has been making strong noises about sharing of publicly funded resources. [We] now have to write*

---

[62] Robert M. Sauer, 'Why Develop Open-Source Software? The Role of Non-Pecuniary Benefits, Monetary Rewards, and Open-Source Licence Type' (2007) 23(4) *Oxford Review of Economic Policy* 605–619 607; Johan Söderberg, 'Free Space Optics in the Czech Wireless Community: Shedding Some Light on the Role of Normativity for User-Initiated Innovations' (2011) 36(4) *Science, Technology, & Human Values* 423–450 430.

[63] This phenomenon has been reported previously (Carol A. Heimer, 'Competing Institutions: Law, Medicine, and Family in Neonatal Intensive Care' (1999) 33(1) *Law & Society Review* 17–66 32).

statements alongside our research proposals explaining how we will manage our data, [which] encompasses software. But I think you raise an interesting point in that researchers are getting mixed messages about wanting to receive commercial impact and value to the national purse, but at the same time [are being encouraged to provide] open access to data and software. I think researchers are pretty confused by this - EU researcher.

A final factor that influenced the development of choice rules as identified by interviewees was the impact of human research ethics requirements as overarching constitutional rules governing the decision to release bioinformatics projects openly. For the majority of interviewees the pure economic rationale for releasing bioinformatics software was straightforward. However, the need to anonymise human research data to comply with data protection legislation was a potential prohibition on the release of data. The *Health Insurance Portability and Accountability Act 1996* requires covered entities (that is, entities processing personal health data) to remove a number of identifiers before release of personal health data. These include any unique identifying number, character or code. Further, the EU General Data Protection Regulations (GDPR) imposes an even higher standard for anonymisation, requiring that re-identification of data, including sensitive categories of personal health data, be rendered impossible. Whilst the Australian *Privacy Act 1988* (Cth) does not impose the same high standard, data privacy and protection legislation remains a strong militating factor against the open source bioinformatics software.[64] Given the importance of research data protection as a safeguard for research participants, the cost of this compliance was not trivial.

The effect of privacy laws on the sharing of biomedical data is complicated by the fact that human genomic data is required to verify the results that were being produced using the software.[65] Accordingly, one US interviewee mentioned that for human genomics analysis software, their project developed a system of using randomised values. In the alternative, they had specific consent to use for those purposes, in accordance with the requirement that participants in research give specific consent for each specific use of their data. Nevertheless, researchers explained the difficulty of obtaining sufficiently anonymised data:

[Basically] we have put a lot of effort into ensuring that nay data that we use in this test suite is either public domain or if it is real human data it has been published with levels of consent that allows the distribution of the suite. The other option is

---

[64] Christine L. Borgman, 'The Conundrum of Sharing Research Data' (2012) 63(6) *Journal of the American Society for Information Science and Technology* 1059–1078 1072.

[65] Jonathan M. J. Derry et al., 'Developing Predictive Molecular Maps of Human Disease Through Community-Based Modeling' (2012) 44(2) *Nature Genetics* 127–130 127-9.

the generation of synthetic data, where it has been completely either synthetically generated or completely de-identified at the atomic level... [This is important] in support for the community, because we like to provide fully runnable examples for tutorials, but for a few of our methods we can't provide these examples because we don't currently have any data which is sufficiently consented. - US Researcher.

Although a lack of consented data was not a bar on the release of open source software per se, it nevertheless had the potential to affect how widely the software was used, which will be discussed in more detail in Section 6.3.1.7 with respect to scope rules.

6.3.1.4 Information Rules

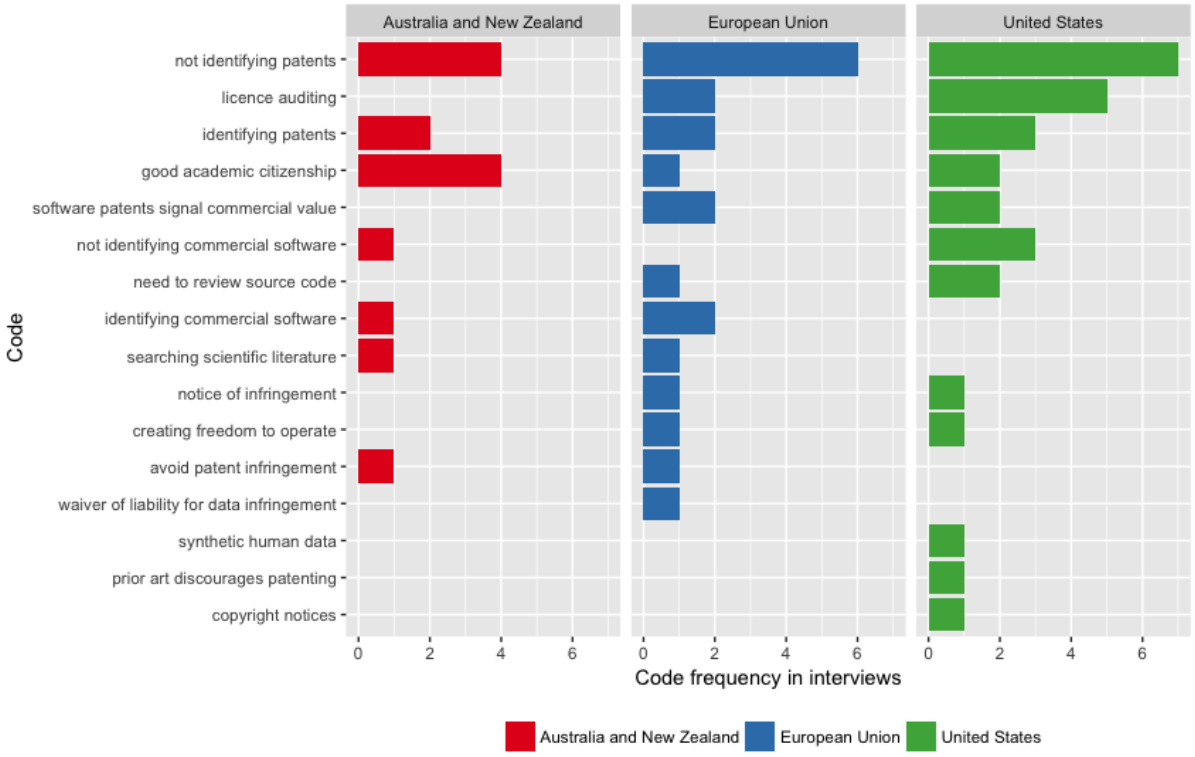

Figure 6.4: Information rules identified through interview transcription.

One anticipated behaviour that emerged from all interviewee transcripts was publications being included in literature searches when researchers were conducting further research in a particular field. Although these literature searches were not explicitly described by interviewees as prior art searches, they nevertheless allowed researchers to identify other researchers who were attempting to solve the same problem. This pattern of behaviour confirms Ouellette's observations that a substantial number of academic scientific researchers search for software as part of literature reviews.[66] However, only one interviewee in the

[66] Lisa Larrimore Ouellette, 'Who Reads Patents?' (2017) 35(5) *Nature Biotechnology* 421–424 421-2.

present study described patent search strategies being mandated as a rule. All other interviewees described this institutional arrangement either as a shared strategy or noted that they had no rules surrounding this requirement. The responsibility of patent searching was instead left to the technology transfer office. The rationalisations for refusing to search for patents were varied. At least three US interviewees expressed concern about the heightened damages for wilful infringement under US patent law:

> *I've always been told that you shouldn't look, because if you get caught it's treble damages because you've been aware of the potential for infringement. So my policy has always been not to look for software patents in the area. And fortunately the area that I work in involves [basic research], there's not a hell of a lot of innovation that goes on or stuff that is patentable.* - US developer

Another US developer expressed scepticism at the value of searching patents for technical information. This perspective was in spite of the existence of key patents in the field which were filed by a private research group using software that had been developed in a public private collaboration:

> *No, that's a horrible idea! Did I read <company> patents at the time when they were released? Absolutely. But here's the thing. We all have to make choices about these things. Because I made this choice to always be working for the public good, I feel like I have the high ground in that I am not out to exploit or commercialise the effort per se, and also I think that the software patent system is broken. So I don't really see why I should try to participate or spend time with the patent system.* - US developer

In addition, of the European interviewees, only one interviewee expressed strong concerns about the impact of the *sui generis* database right. The remainder mentioned that they and their collaborators either ignored the right or included a waiver as a 'third party' transmitter of that data:

> *The status of intellectual property rights around genomic databases is complicated because we are gathering information from many sources. If you follow each of our databases you will find that our data comes from thousands of sources. We manage this by placing a statement on our website that <institute> promises that we place no additional restrictions on data, but at the same time we are not responsible for any third party rights that exist on that data. [In this regard] we have been very principled, and we have refused any click through recognition... [That being said],*

it's still a regular question (and by regular I mean two or four times a year) when people ask us for the full set of restrictions on a dataset. [That's why we have] the disclaimer of liability. - European researcher.

In essence, this approach to data accumulation almost operates as an informal clearinghouse norm to protect bioinformaticians from being sued for copyright infringement by engaging with the first author to get access to the data.[67] Other interviewees mentioned that they were much more heavily reliant on formal instruments for the exchange of data, and for that reason expressed a preference for public domain data:

There have been some changes in intellectual property status for bioinformatics databases where the ownership has changed. The most famous would be [database] at [institute], where the government funding for the project dried up so the project leads sought to licence it with a new type of licence. Similarly, at [institute] there was a change in licence for [database 2] and [database 3], which were both hosted in the US. And that was motivated by similar reasons, due to a lack of funding, so in all three cases they've left - European researcher

This question of reciprocal licensing for data exchange neatly dovetails into the next section of this chapter which explores evidence garnered on to payoff rules.

6.3.1.5 Payoff Rules

Under this category, payoff rules identified in the present study related to the benefits that flowed from the use of open source software, as well as the costs that had to be paid for use (whether these costs were licensing fees or attribution). These rules could include rules that were imposed by open source software development as well as rules from university and research institute technology transfer offices.[68] Crucially, the graph above demonstrates a difference in payoff rules between jurisdictions, with US interviewees describing significantly more potential benefits from the development of bioinformatics software. Further, technology transfer policies at different research institutes, as well as individual perspectives on patent strategy, were described by researchers as having a greater impact on propensity to patent.

---

[67] Melanie Dulong de Rosnay and Andrés Guadamuz, 'Open Access to Biodiversity Scientific Data: A Comparative Study' (Paper presented at *17th International Consortium on Applied Bioeconomy Research ICABR Conference on Innovation and the Policy for the Bioeconomy* , June 2013) 15.

[68] Joachim Henkel, 'Selective Revealing in Open Innovation Processes: The Case of Embedded Linux' (2006) 35(7) *Research policy* 953–969 960.

Figure 6.5: Payoff rules identified through interview transcription.

In addition, payoff rules in the present study were also inductively coded in the context of sanctions for breach of open source licences. One key finding of this thesis is the almost complete absence of any litigation or informal enforcement of rights between interviewees. Although bioinformatics patent disputes are relatively rare, as Section 4.2.2 of Chapter Four discussed, with respect to university software patents the situation (at least in the US) is somewhat more ambiguous. The Wisconsin Alumni Research Foundation (WARF) has aggressively pursued software patent infringement action against Intel[69] and Apple[70] in the US District Court of Wisconsin, forcing settlements in each case. In a separate case, Carnegie Mellon University won an estimated 1.54 billion US dollars in patent infringement remedies against Marvell Technology Group for the infringement of two patents with respect to hard disk drive algorithms[71]. Love has described these cases as 'watershed moment[s]' for University software patent infringement.[72] Despite the massive awards of damages in these cases, Chien notes that non-profit entities only account for one percent of patent infringement

[69] *Wisconsin Alumni Research Foundation v Intel Corporation* 656 F.Supp.2d 898 (2009).

[70] *Wisconsin Alumni Research Foundation v Apple* 135 F. Supp. 3d. 865 (2015).

[71] *Carnegie Mellon University v Marvell Technology Group Ltd and Marvell Semiconductor Inc* 807 F.3d 1283 (2015).

[72] Brian J. Love, 'Do University Patents Pay Off-Evidence from a Survey of University Inventors in Computer Science and Electrical Engineering' (2013) 16(2) *Yale Journal of Law and Technology* 285 291.

plaintiffs in litigation in the US.[73] Only two US interviewees mentioned that they had been involved in disputes over bioinformatics software. One mentioned that they had been involved as an expert witness in a patent trial concerning a non-bioinformatics patent. The other referred to a dispute over commercially licensed bioinformatics software that was solved through in house negotiation between the two research universities. This lack of litigation suggests that bioinformatics patents are regarded as too difficult or not valuable enough for universities to enforce.

Another, perhaps more unexpected, behaviour that was reported from one interview (from an European researcher) was the reference to trade marks as a means of protecting open source bioinformatics software. Legal literature has discussed how trade marks have been used to protect open source projects.[74] However, this study is the first to report trademarks being used for open source software in bioinformatics. In particular, this interviewee discussed how the use of trademarks was often used to protect the reputation of the institution for advertising:

> *For example, <project> name has been trademarked, and we wouldn't want anyone using our trademarked software name in their licensing campaigns to imply that they have been endorsed by us. And there have been instances where we have contacted these third parties, informally, to tell them to stop using our software name, but we've done so in an informal way.* - European researcher

This quote demonstrates the importance of recognition within scientific software development. However, the reliance on trademark protection indicates one of the key inhibiting factors of acquiring patent protection for an open source project, namely, the cost of patent searching, filing and upkeep. As demonstrated by the quantitative analysis in Chapter Five, the number of either pure basic or use inspired basic research patents is relatively low compared to patents filed and obtained for computer science related inventions.[75]

6.3.1.6 Aggregation Rules

Most of the aggregation rules identified through inductive coding in the present study related to either collaborative development strategies, collective decisions over relicensing and the formation of standards. As discussed in Section 6.3.1.2, the vast majority of interviewees did

---

[73] Colleen V. Chien, 'Of Trolls, Davids, Goliaths, and Kings: Narratives and Evidence in the Litigation of High-Tech Patents Frontiers in Empirical Patent Law Scholarship' (2008) 87(5) *North Carolina Law Review* 1571–1616 1600.

[74] Heather Meeker and Stephanie Petit, 'The New Foundations of Open Source' (2016) 33(3) *Santa Clara High Technology Law Journal* 415–426 420-1.

[75] Arti K. Rai, John R. Allison and Bhaven N. Sampat, above n27, 1534.

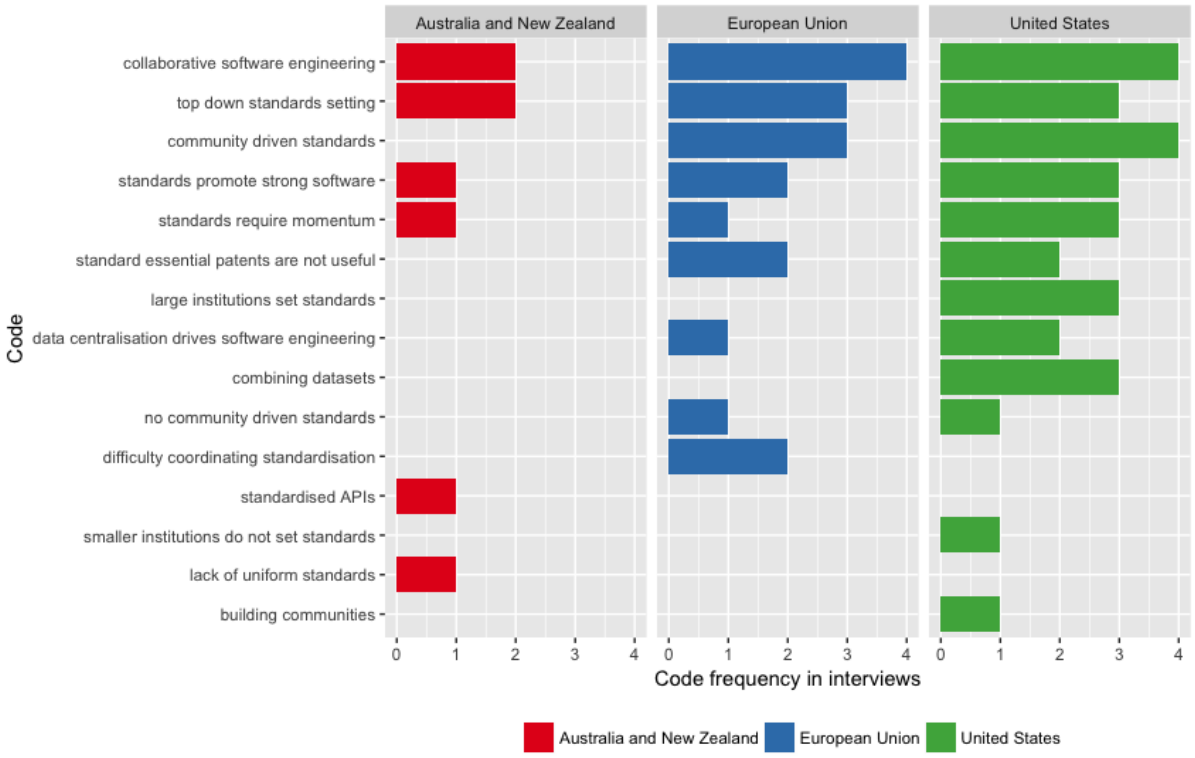

Figure 6.6: Aggregation rules identified through interview transcription.

not describe rules for handling external contributions to their projects. However, a minority (five US interviewees, three EU interviewees and one Australian interviewee) discussed norms for recognising the contributions of external developers. Two of the US interviewees described these as project specific standards, where developers would receive authorship and guidance over choices of development depending on their contributions. The remaining five interviewees described the same norms but noted that these shared strategies operated on a largely informal basis, where researchers would contact one another directly to work together on a particular project.

> *I have had situations where - not on purpose but you have discovered that you have worked on something similar to another researcher at another institute. In these circumstances, you come to an agreement where one group will publish first and another group will publish later will providing attributions to the first group. Another situation that comes to mind is where we working on a particular project and we had a colleague who had worked with my supervisor previously and thought that she may have had an interest on the same sort of work. I contacted her to ask if she was interested in collaborating* - Australian researcher.

These informal shared strategies can be contrasted with genomic data access agreements,

which require institutional approval.[76] Likewise, the formation of standardised APIs and file formats was a key concern for almost all interviewees. Unlike other fields of software engineering, where patent pledges have emerged, no interviewees reported using patents for the formation of standards. To quote one European interviewee, who had acquired a valuable patent on a novel bioinformatics technique:

> *I don't think we've established a standard... as long as other people want to use those, we've got a right to do that. But that's not to say that everyone else has to follow our lead. The subject is not quite ready for standards yet. The main reason for acquiring the patent [and licensing it to a company] was because we deemed the techniques disclosed valuable* - European researcher.

However, an Australian interviewee, whilst noting that the vast majority of bioinformatics software was used for pure or basic inspired research only, described the potential for standardisation of bioinformatics as follows:

> *That would probably be useful in some cases. Some colleagues of mine doing public health work that are doing similar research to what I am doing [in a pure use inspired context]. Again, that's an environment where there is more regulation and I can see how this could encourage the development of more robust supported software that would integrate these different bioinformatics methods in a more well developed framework* - Australian researcher.

Nevertheless, the majority of interviewees when questioned mentioned that neither their projects nor their institutions had organisational rules in place for the creation and enforcement bioinformatics standards. An issue identified by one European interviewee was the difficulty in coordinating standards formation, particularly from a structured top down approach:

> *There's still a lot of smaller organisations though, and I think that is a problem with some of the larger standards setting organisations. They're very top down and very structured, and there's a lot of dragooning researchers into using a particular format. For a lot of bioinformaticians who are buried in a biological organisation, there's probably a lot of difficulty in justifying attending all the meetings to your superiors. Some of my colleagues participate in some of these meetings in almost a clandestine fashion, and as bioinformatics and ecoinformatics professionals we*

[76] Tempest A. van Schaik et al., 'The need to redefine genomic data sharing: A focus on data accessibility' (2014) 3(4) *Applied & Translational Genomics* 100–104 101.

need to participate in these things, despite the fact they are outside our usual role. The exception is where there's a very visible public project where data sharing is so core to its activity that they understand why you need to do it. My director has tasked me with participating in a particular research consortium, and I think they are doing a good job promoting standards in this field - EU researcher.

Despite these large scale initiatives, observation suggests that for the most part, researchers involved in bioinformatics projects did not consider the possibility of standards development when conducting research, and there were no institutional arrangements for standardisation of bioinformatics software.

6.3.1.7 Scope Rules

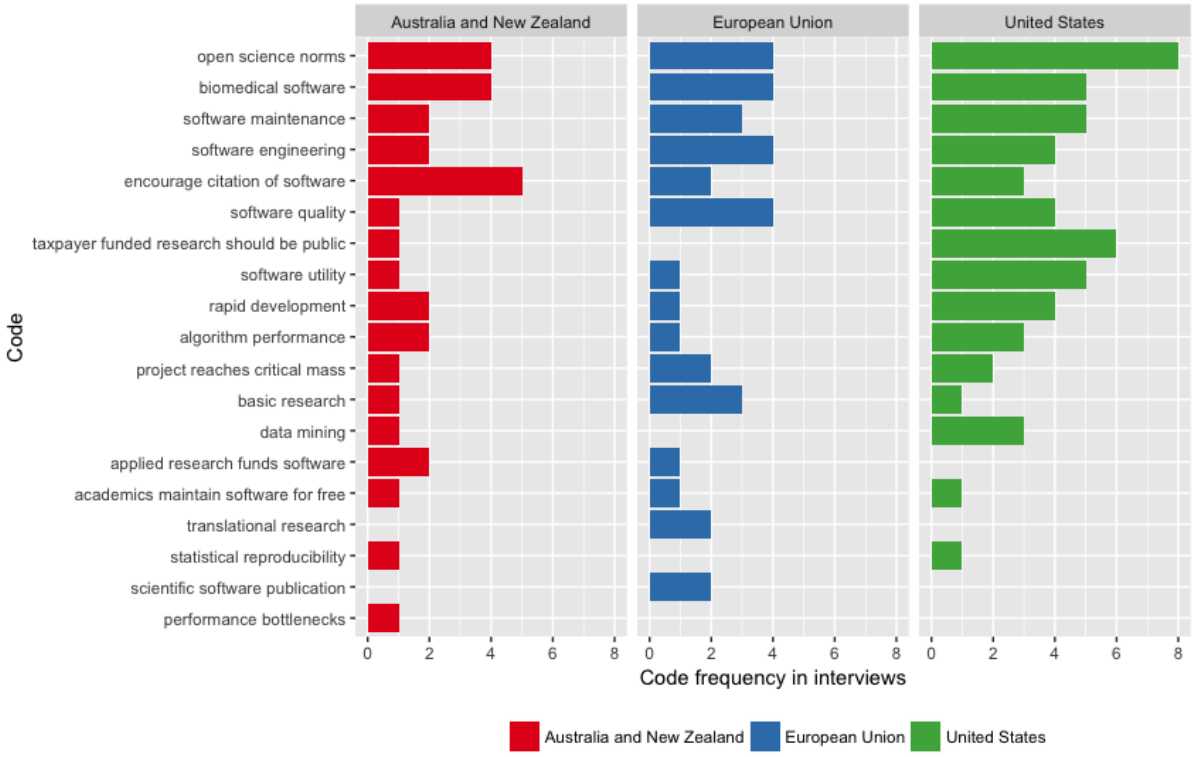

Figure 6.7: Scope rules identified through interview transcription.

The scope rules identified through inductive coding in the present study related to the reasons for writing scientific software, as well as the objectives of each research in releasing their software under an open source licence or attempting to commercialise their work. As articulated earlier when discussing the different classes of bioinformatics contributors, there was wide divergence amongst interviewees of measures of success and failure. One of the key measures that was identified by different researchers was the notion of encouraging wide use and reuse of software:

> *Getting the software out there and lowering the barriers to entry is the main thing… (Particular subfields of bioinformatics) are arguably tiny markets, and the community worldwide for each specific area of research. If I have a thousand users, I consider that to be a success. So I think the issue of open source versus closed source… I think open source is a better way to get people to use the software… To acquire patents would keep people from using that software* - US Researcher

However, one Australian interviewee discussed how patents provided them with an avenue to receive research recognition without having to rely on publishing within scientific journals. This interviewee attributed their preference to the fact that highly ranked scientific journals might be biased towards reporting on the results of a particular research experiment performed with an algorithm. By failing to focus on the algorithm, this effect would lead to a lower degree of recognition for the software developer as opposed to the designer of the experiment. This lower degree of recognition then impacted future funding for other research projects. Moreover, interviewees across all three jurisdictions and technical backgrounds discussed the difficulty in obtaining funding for software maintenance once the main research project had been completed:

> *[One of the major factors] in the success and failure [of open source bioinformatics projects] have been the policies of funding agencies. So there might be two or three renewals of funding, but after that you are basically in the valley of death situation where you don't have enough adopters to support the software. So you have to look at other ways of getting it funded, and possibly looking at a commercial licensing model. But very few people find that's a viable funding model.* - European researcher.

Two other Australian and New Zealand interviewees discussed the benefit of altmetrics as a source of recognition. These include mentions of software through social media as well as the number of contributions and downloads of a particular software package. In addition, three US interviewees mentioned that they had attempted to build software download rates as a measure of success into their funding applications, and there is some empirical evidence to suggest that altmetrics may indicate whether an article has high research impact:[77]

[77] Rodrigo Costas, Zohreh Zahedi and Paul Wouters, 'Do "Altmetrics" Correlate with Citations? Extensive Comparison of Altmetric Indicators with Citations From a Multidisciplinary Perspective' (2015) 66(10) *Journal of the Association for Information Science and Technology* 2003–2019 2013-4.

> *I would say that you use BSD because I am a public employee and I want to maximise the downstream use of my software. I know that in the past <company> has taken pieces of my software for distributed analysis pipelines and I was totally okay with that. They probably wouldn't have made that decision to reuse the software if they had been dealing with GPL licensed software. That being said, the majority of the software I develop are algorithms and data structures that are used by the data science community. The data science community has settled on BSD compatible licences, because so much of that research is done in partnership with different companies.* - US Researcher

The most common way to enhance academic reputation was through traditional academic publications. Patent protection was rarely described by interviewees as an optimal solution to solving issues of long term maintenance for scientific software due to the high reputation risk associated with patenting a piece of community developed software.

6.4 LIMITATIONS AND DISCUSSION OF RESULTS

6.4.1 *Limitations*

The results presented in this chapter suffer by their very design from two key limitations. First, as discussed in the institutional analysis literature, the IGT metric is frequently used to analyse particular case studies of common pool resources.[78] These case studies usually involve considering the entire set of rules that are involved in the management of that commons resource. Further, as discussed frequently within the grounded theory literature on open source software, the role that intellectual property rights play in the success or failure of a project may be only one factor for determining whether the project is successful or not. For example, the research of Charles Schweik suggests that licence type is not a statistically significant factor in determining whether a particular open source project will be subject to ongoing use and development.[79] In applying the IGT metric solely to study the role that patents and copyright protection play in the management of open source bioinformatics projects, it is possible that this study avoided examining a host of other structural issues that could be responsible for the success and failure of these projects. However, as this thesis is a legal thesis, adopting such a level of analysis would fall outside its scope.

A second limitation of this study is that the interview results were entirely collected from participants who self selected for this study. Although the sampling strategy for this stage of

---

[78] Saba Siddiki, Xavier Basurto and Christopher M. Weible, above n51, 173-4.

[79] Charles M. Schweik and Robert C. English, above n20, 173.

empirical research was specifically designed to capture information from bioinformatics researchers who had released open source software and had either filed for their own patents or were involved in institutions that had acquired patents. Given the cultural norm towards openness described by Stodden and others, it is possible that the respondents for this study did not accurately report how they viewed the utility of software patents.[80] Nevertheless, the fact that many interviewees did discuss their interactions with the patent system (either as patent applicants, cross licensees of patented technology or users of patented software) as well as copyright protection does lend validity to the accuracy of the results presented. Finally, the purpose of qualitative research is not to necessarily draw causal conclusions about particular observations. Instead, the purpose of qualitative research is to develop an understanding of how interviewees perceive particular events. Accordingly, these results cannot be taken as suggesting any specific causal inferences. However, they do demonstrate how bioinformaticians perceive the intersection between formal intellectual property rights and user generated norms in open source bioinformatics.

6.4.2 *Benefits and Discussion of Results*

6.4.2.1 Deductive Coding Results

The IGT coding metric presented in this Chapter represents a useful mechanism to understand how different open source bioinformatics projects are structured around patent and copyright protection. Crucially, no significant differences were identified in the classification of different institutional arrangements by jurisdiction. This result may suggest that copyright and patent strategies are not influenced by the jurisdiction where researchers are located. Instead, patent and copyright strategies are influenced by whether researchers are prepared to conduct commercialisation in other jurisdictions. For example, despite explicitly referring to commercialising publicly funded research as 'double taxation', two interviewees in Australia and New Zealand respectively were prepared to patent open source software in the US. These interviewees saw patenting a means of providing their with defensibility work from appropriation (despite each interviewee expressing their misgivings as to the effectiveness of this approach).

In addition, the IGT coding metric is useful as a diagnostic model for understanding the shortcomings of different commons based arrangements by observing how frequently or infrequently different categories of rules emerge. The fewest number of institutional arrangements identified were aggregation rules; institutional arrangements that govern the

[80] Victoria Stodden, above n28 26 <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1550193>; Victoria Stodden, Peixuan Guo and Zhaokun Ma, 'Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals' (2013) 8(6) *PLOS ONE* e67111 7.

decision about whether to acquire patent protection or to commercialise a piece of bioinformatics software. Given that commons based approaches are dependent on the co-operation of stakeholders within a particular project, this observation suggests open source development could be undermined. In particular, it may suggest that despite the presence of open source licensing, certain project contributors may be overruled or ignored when it comes to acquiring patents or commercialising software. At least one interviewee described how the acquisition of a patent without adequate consultation with the remainder of the development team led to a decrease in the use of the software. The absence of institutional arrangements for the development of standards has led to a fragmentation of standards across different open source projects at different institutions. In addition, the absence of standards may explain in part why the Global Alliance for Genomics and Health (GA4GH) has attempted to develop a standardised series of pipelines for the exchange of genomic data in human genomics research.[81]

The second fewest number of institutional arrangements that were identified were for information rules. In other words, there was a consistent absence of institutional arrangements for either identifying patents or commercial software. That is, when asked, the majority of interviewees stated that they did not search for patents or commercial licences on software. Where these institutional arrangements were identified, they were shared strategies that had been adopted at the initiative of individual researchers and therefore had no penalties or deontic components. An absence of referral to the patent literature could be demonstrative of the relative lack of value of patents for bioinformaticians. In the alternative, it could also demonstrate a reliance on technology transfer offices to perform freedom to operate searches. Finally, it could reflect the presence of a social norm reinforcing the research exemptions in EU, Australian and New Zealand patent law. However, the internationalisation of scientific technology transfer could create problems in jurisdictions where no such research exemption exists (namely the US). In addition, there was a relative absence of explicit institutional arrangements for avoiding using incompatibly licensed software (such as using GPL licensed software packages in a permissively licensed project). Although one institution had very clearly defined policies for determining when such software can be used in concert, an absence of these institutional arrangements in other institutions could lead to disastrous consequences with respect to software copyright infringement.[82] Finally, there were noticeably more payoff rules identified by US interviewees. However, this finding could merely indicate that US researchers are more conscious of intellectual property issues than their counterparts in Europe, Australia and New Zealand.

---

[81] Lillian L. Siu et al., 'Facilitating a Culture of Responsible and Effective Sharing of Cancer Genome Data' (2016) 22(5) *Nature medicine* 464–471 467.

[82] Andrew Morin, Jennifer Urban and Piotr Sliz, 'A Quick Guide to Software Licensing for the Scientist-Programmer' (2012) 8(7) *PLoS Comput Biol* e1002598 3-4.

6.4.2.2 Inductive Coding Results

The inductive coding of rule types also revealed significant information about the formation of rules used to govern different bioinformatics projects, as well as the role that copyright and patent law played in the formation of these rules. In particular, these interviews revealed that there are a number of different types of open source project within bioinformatics research. The bulk of open source bioinformatics projects start as small scale projects that are developed inside a single academic institute by a sole developer or team of developers. Whilst these projects may receive initial funding as part of a project grant, there is unlikely to be any follow on funding for the maintenance of software due to the small potential development base.[83] From this starting point, there are a number of possible trajectories that an open source bioinformatics project may take. One possible trajectory, the 'anti-commons' trajectory described by Schweik and English, involves the project falling into disuse due to a lack of funding or support. At least one European institute had developed a system for adopting abandoned software packages and supporting them to avoid this scenario. Further, if a project has significant utility for bioinformatics researchers, it may achieve sufficient importance to be developed on an ongoing basis or a cross institutional scale. In these circumstances, the project team members may be able to apply for funding for ongoing maintenance. Finally, if the project has sufficient utility, it could be cross licensed, or bundled as part of a commercial service.

What role do copyright and patent protection play in each of these development models? There will always be a certain percentage of open source bioinformatics projects that have limited general use or will be abandoned by their developers due to the transient nature of academic research. However, attribution is a key motivator for academic bioinformatics developers. This attribution may occur through formal mechanisms, such as citation in academic outputs, or informal mechanisms, such as community engagement with the software. Accordingly, contribution agreements and what several interviewees described as 'good academic citizenship' can, in concert with copyright protection, ensure that bioinformaticians receive appropriate attribution for their work. However, one universal theme that emerged was that copyright was perceived as inhibiting open source bioinformatics software where there was a lack of clarity regarding copyright ownership and authorship. For example, two US interviewees mentioned Docker and Juypter, two popular server administration toolkits for running a virtual operating system within another operating system to perform a specific service.[84] Although each of these software tools are released under open

---

[83] Charles M. Schweik and Robert C. English, above n20, 43.

[84] Björn A. Grüning et al., 'Jupyter and Galaxy: Easing Entry Barriers into Complex Data Analyses for Biomedical Researchers' (2017) 13(5) *PLOS Computational Biology* e1005425 2.

source licences, both interviewees mentioned instances where proprietary software was bundled with a publicly released Docker package. These examples demonstrate that successful open source bioinformatics projects need careful code auditing strategies to avoid copyright infringement.

With respect to patent acquisition, the reverse effect was observed. In other words, patents were perceived as inhibiting academic reputation by obfuscating whether the patented software was actually available to be reused. This finding appears to contradict the findings from Section 5.3.2 of Chapter Five suggesting grant of a patent did not have an impact on the citation rates associated with patent publication pairsȦ possible explanation for this phenomenon though is that, as discussed in Section 5.3.1, the citation counts were highly skewed, so that certain bioinformatics software packages received many more citations than others. This trend that has been reported in other studies into bioinformatics database reuse.[85] Moreover, patent holding interviewees also identified non reputational reasons to acquire patents, such as the need to signal commercial value of investors or, more interestingly, the need to develop robustly engineered software to satisfy regulatory requirements for biomedical software. In particular, the latter rationale demonstrates that there is at least one strong technical rationalisation to only partially release source code.[86] However, because of the negative effect on reputation of patent acquisition within the open source community, most interviewees reported that they did not engage with the patent system. These interviewees instead relied on 'software as a service' models, including designing bespoke software and performing analysis for other researchers. Nevertheless, the interviewees who relied on these licensing models still placed high importance on the need to ensure that these changes were contributed back into the main open source project. These strategies to guide sustainability will be discussed in further detail in Chapter Seven.

6.5 CONCLUSION

This Chapter describes a grounded theory methodology for analysing the institutional arrangements that are used to govern open source bioinformatics projects. It also describes the Institutional Grammar Tool model that was used to assess the prevalence of different institutional rule categories. In particular, the grounded theory from the interviews suggests that open source development practices are heavily embedded in bioinformatics research across all jurisdictions under examination. This grounded theory also suggests that patents are not perceived as have a significant inhibitory effect on bioinformatics research except in isolated cases. However, almost no interviewees described patents as having a positive impact

---

[85] Geraint Duck et al., 'A Survey of Bioinformatics Database and Software Usage through Mining the Literature' (2016) 11(6) *PLOS ONE* e0157989 8-9.

[86] Joachim Henkel, above n68 960.

on the formation of rules for the governance of open source bioinformatics communities, irrespective of jurisdiction. Interviewees in the US reported more institutional pressure to commercialise (which is consistent with the effects of the *Bayh-Dole Act*) and more payoff institutional statements. However, patenting was not a significant motivating factor in the development of academic computational biology software. The results suggest that across jurisdictional boundaries there is an absence of rules for collective decision making regarding patent protection and technology transfer for bioinformatics software. Further, there is an absence of informational institutional arrangements for identifying bioinformatics patents and conflicting licences. The next chapter accordingly provides recommendations on best practices that can be used to improve the uptake of these rules.

Chapter 7

PRIVATE ORDERING & LEGISLATIVE STRATEGIES TO PROMOTE OPEN INNOVATION IN COMPUTATIONAL BIOLOGY RESEARCH

7.1 INTRODUCTION

Chapters Five and Six provided an analysis of the results garnered from patent landscaping and semi-structured interviews with bioinformaticians who had been involved in open source software licensing. The purpose of this empirical research was twofold. The first reason was to measure whether academic computational biologists experienced any positive or negative effects flowing from patent and copyright protection when developing open source bioinformatics software. The second reason was to determine whether these effects are influenced by jurisdictional divergences in copyright and patent protection. These empirical results were framed within the context of the doctrinal and theoretical analysis of both formal intellectual property laws and informal norms for commons based development in Chapters Two, Three and Four. The aim of this chapter is to analyse what private ordering and funding strategies might be needed to improve the interface between open source development in bioinformatics and intellectual property protection. This chapter focuses primarily on Australia, but is informed by the broader doctrinal and empirical analyses that have been undertaken in the US, the EU and New Zealand, as well as in Australia.

In addition, it explores a set of recommendations on legislative strategies for copyright and patent reform. These legislative strategies will primarily focus on increasing the consistency between comparative copyright and patent legislation as it applies to software. This call for legislative reform reflects the policy pluralism around publicly funded scientific research where intellectual property and non intellectual property incentives are used to guide research. Commons scholars have traditionally argued against using the Institutional Analysis and Development (IAD) framework as a means to guide normative intellectual property reform. This caution arises from the difference between the resource types within a natural resource commons and a knowledge commons.[1] However, these mechanisms for law reform are concerned less with the resources under governance (insofar that patent and copyright laws are designed to be technologically neutral). Instead, these mechanisms are more concerned with how the holder of these property rights can use them to enforce rules for governing an open source project.[2] In addition, open source bioinformatics projects may operate on both a national and an international scale. There is therefore a strong incentive to confirm that open source licences operate similarly in Australia and New Zealand as they do in the US or the EU.

---

[1] Daniel H. Cole, 'Learning from Lin: Lessons and Cautions from the Natural Commons for the Knowledge Commons' in Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press 2014)45–68 48-9.

[2] Helena Howe, Jonathan Griffiths and S?verine Dusollier, 'The Commons as a Reverse Intellectual Property – From Exclusivity to Inclusivity' in *Concepts of Property in Intellectual Property Law* (Cambridge University Press 2013)258 266-7.

This chapter is split into three parts. Section 7.2 explains how there may be different intellectual property management strategies may be needed for bioinformatics research in Australia compared with the US and the EU and, to a lesser extent, New Zealand. Section 7.3 then discusses different private ordering strategies for computational biology software development, flowing from the analysis of the semi-structured interviews contained in Chapter Six. These strategies focus on the development of institutional arrangements, including university technology transfer mechanisms and innovative funding models, that can support open source development in bioinformatics research. Section 7.3 concludes by examining the limitations of these private ordering strategies. Section 7.4 discusses potential legislative strategies that can reduce the transaction costs associated with copyright and patent allocation mechanisms in computational biology. These recommendations are largely targeted at Australian copyright and patent law to balance consistency in other jurisdictions with economic suitability. The final recommendation is a speculative suggestion for a new form of intellectual property that could be used to protect the functional aspects of computer programs whilst supporting open source development.

7.2 THE ROLE OF BIOINFORMATICS IN THE AUSTRALIAN INNOVATION LANDSCAPE

Although the private ordering strategies discussed below in Section 7.3 are not constrained to any one jurisdiction, the legislative strategies discussed in Section 7.4 are largely focused on patent and copyright reform in Australia. The vast majority of legal literature on the differing roles of intellectual property rights in software development across jurisdictions has contrasted either the US and the EU or the US, the EU and Japan. The former approaches have largely focused on the differences between US and EU jurisprudence on copyright and patent protection. These differences include the extent of functional protection under copyright law for software, the scope of patent protection for intangible inventions or the impact of legislative reform such as the *Bayh-Dole Act* on academic research.[3] The latter approach is often referred to as a 'trilateral' or 'triadic' approach to studying intellectual property (particularly patents) due to the equivalent economic activity of each jurisdiction.[4]

However, the Australian economy is significantly smaller than the 'trilateral' economies

---

[3] Christoph Laub, 'Software Patenting: Legal Standards in Europe and the US in view of Strategic Limitations of the IP Systems' (2006) 9(3) *The Journal of World Intellectual Property* 344–372 362-6; Jerome H. Reichman and Ruth L. Okediji, 'When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale' (2012) 96(4) *Minnesota Law Review* 1362–1480 1378-84; Geertrui Van Overwalle, 'Governing Genomic Data: Plea for an "Open Commons"' in Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press 2014)137–154 151.

[4] Paola Criscuolo, 'The 'Home Advantage' Effect and Patent Families. A Comparison of OECD Triadic Patents, the USPTO and the EPO' (2006) 66(1) *Scientometrics* 23–41.

(particularly with respect to information communications and technology (ICT) research) and can be best described as a 'small market economy'.[5] This economic disparity invites a discussion as to whether success in publicly funded research in these economies is best achieved through government intervention, contractual negotiation between parties or intellectual property reform.[6] On the one hand, the analysis of the semi-structured interviews in Chapter Six suggests that where researchers are located does not influence the normative rules that govern open source computational biology projects. On the other hand, interviewees mentioned that they adopted different strategies depending on where they were conducting their research. Further, jurisdictional divergences in what actions are permitted *outside* of copyright and patent under fair dealing and research exceptions may influence how commons resources are governed.[7] Finally, jurisdictional differences in public funding strategies and technology transfer was found to have the potential to influence how different projects are governed. In particular, increased institutional pressure to commercialise (as was observed amongst US interviewees) was a key difference between the jurisdictions observed.

An additional factor is the impact of national innovation strategies on open source software development. This field has been somewhat uncharted due to the fact that the majority of open innovation studies focus on firm activity.[8] By contrast, national innovation strategies examine how innovation occurs in a country. This examination may extend to the performance of both private and public industries, as well as funding of different research fields.[9] Annette McLeod's comprehensive overview of funding for Information Communications Technology (ICT) research in Australia notes that funding of bioinformatics research is guided by the Australia and New Zealand Standard Research Classification (ANZSRC) scheme. Under this funding scheme, governments and non-profits spent 407 million Australian dollars, or 40 percent of their total research expenditure, on ICT research.[10]

However, ICT research which supports other scientific research, including bioinformatics research, is not included in the ICT ANZSRC class. Instead, bioinformatics falls under the Molecular Biology and Genomics ANZSRC classifications. This fragmentation was

---

5 Susy Frankel, *Test Tubes for Global Intellectual Property Issues* (Cambridge University Press, 2015) 14-5.

6 Partha Dasgupta and Paul A. David, 'Toward a New Economics of Science' (1994) 23(5) *Research Policy* 487–521 496.

7 Lisa Larrimore Ouellette, 'Patent Experimentalism' (2015) 101(1) *Virginia Law Review* 65–128 72.

8 Although more recent studies have focussed on the role that open innovation plays in national innovation systems (Sarah E. Ali-Khan et al., 'Defining Success in Open Science' (2018) 2(2) *MNI Open Research* 1 4).

9 Jeroen P. J. de Jong, Tarmo Kalvet and Wim Vanhaverbeke, 'Exploring a Theoretical Framework to Structure the Public Policy Implications of Open Innovation' (2010) 22(8) *Technology Analysis & Strategic Management* 877–896 879.

10 Annette McLeod, *Returns on Investment: Considerations on Publicly Funded ICT Research and Impact Assessment* (PhD thesis, University of Melbourne, 2016) 8 <http://minerva-access.unimelb.edu.au/handle/11343/124272>.

recognised in the Australian Bioinformatics Network's most recent final report, which called for greater training and integration of bioinformaticians into research projects.[11] Nevertheless, the fragmentation of the Australian bioinformatics community has been compounded by the shift in Australian research technology transfer policies away from a centralised national innovation system dominated by CSIRO. Technology transfer is now largely driven by individual universities, which have attempted to replicate the operation of the US *Bayh-Dole* Act.[12]

However, attempts to overlay these strategies onto bioinformatics are complicated by the heterogenous, fragmented nature of the community in Australia and the differences in appropriate technology transfer models for biotechnology and computer science research.[13] Accordingly, this thesis posits that publicly funded bioinformatics development should be tied to technology transfer strategies that encourage the maximum diffusion and reuse of bioinformatics software.[14] The outcome of the mixed methods approach in this thesis has been to demonstrate that this diffusion is best achieved through open source licensing.[15] This finding is also reinforced by the existence of carefully defined choice rules identified through the interview stage of this study.[16] These choice rules indicated that interviewees are careful in determining how an open source software package was best licensed beyond altruistic considerations about open culture. These strategies demonstrate that bioinformaticians employ rational cost/benefit considerations in releasing software under an open source licence.

Open source licensing was identified by interviewees as an appropriate alternative to patenting, because bioinformatics software was 'upstream' research (that is, basic research in a laboratory). However, that research could lead to the production of 'downstream research' (or use inspired research) potentially eligible for patent protection, such as pharmaceutical products or synthetic biology research. These observations demonstrate that not every piece of bioinformatics software needs to be immediately capable of commercial application to

---

[11] Biotechnology Australia, *Australian Bioinformatics Network Final Report* (2009) 5-6 <http : / / australianbioinformatics . net / storage / downloads / Australian % 20Bioinformatics % 20Network % 20Project % 20 - %20Final%20Report%20 - %2020080118.pdf>.

[12] Garrett Upstill and David Symington, 'Technology Transfer and the Creation of Companies: the CSIRO Experience' (2002) 32(3) *R&D Management* 233–239 234; David C. Mowery and Bhaven N. Sampat, 'The Bayh-Dole Act of 1980 and University–Industry Technology Transfer: A Model for Other OECD Governments?' (2004) 30(1-2) *The Journal of Technology Transfer* 115–127 116.

[13] As discussed in Section 6.3.1.1 of Chapter Six when discussing position rules

[14] Commonwealth Secretariat, 'The Role of National Intellectual Property Laws in Promoting Innovation, Scientific and Technological Development' (2017) 43(3-4) *Commonwealth Law Bulletin* 471–488 480.

[15] Sascha Friesike et al., 'Opening Science: Towards an Agenda of Open Science in Academia and Industry' (2015) 40(4) *The Journal of Technology Transfer* 581–601 583-4.

[16] As discussed in Section 6.3.1.3 of Chapter Six.

provide value to its developers. In addition, as indicated by interviewees,[17] the majority of bioinformatics projects will not evolve into either successful open source or commercial projects. Funding agencies must consider this factor when providing funding for bioinformatics research. The private ordering and public funding strategies in Section 7.3 are not generalisable to all bioinformatics projects. However, the benefit of the rich qualitative analysis in Chapter Six is that it can provide some assistance on how to promote the diffusion of publicly funded bioinformatics software both on a domestic and international level.

7.3 PRIVATE ORDERING AND PUBLIC FUNDING STRATEGIES

7.3.1 *Improving Patent and Copyright Notices in Open Source Bioinformatics Projects*

The first private ordering strategy suggested by this study is the introduction of information norms and rules into the governance structure for open source projects. From the institutional statements identified by interviewees,[18] there was a noticeable absence of statements classified as information rules. In particular, there was an absence of rules and norms particularly with respect to patent search strategies across all jurisdictions. In part, this absence can be attributed to the small number of bioinformatics patents potentially impacting academic research outside the US.[19] In addition, the absence of information rules directly relates to the boundary and choice rules identified by inductive coding. Most interviewees identified either norms or shared strategies with respect to licensing strategies for software, unannotated and annotated data, and chemical compounds flowing from bioinformatics development. However, only a minority of interviewees described institutional statements with respect to searching for bioinformatics software patents. This perspective may have been influenced by the fact that the majority of interviewees did not view patents as valuable in the context of bioinformatics development. This perspective was in spite of the fact that each interviewee was listed as an inventor for a bioinformatics related patent or had worked for a research institute or university that had acquired patents.

However, the patent landscaping model described in Chapter Five suggests that an increasing number of research institutes are applying for bioinformatics and computational biology patents. This patent landscaping model does not take into account the bioinformatics patents, as well as patents in related technological fields such as artificial intelligence, that are assigned to private researchers. The consequences for infringing on these patents may be underestimated by academic bioinformatics developers.[20] The need to identify third party

---

[17] See the section on choice rules in Section 6.3.1.3 of Chapter Six

[18] See Section 6.3.1.4 in Chapter Six.

[19] As demonstrated by the patent analysis in Section 5.3.1 of Chapter Five.

[20] Iain M. Cockburn, Rebecca Henderson and Scott Stern, 'The Impact of Artificial Intelligence on Innovation'

intellectual property rights and avoid patent infringement is akin to broader information security policies. These policies are predicated on avoiding the economic consequences of cyber security or data protection breaches.[21] The negative economic externalities that are prevented by an information security policy may include the loss of personal data or damage to computer infrastructure through malware or unauthorised intrusion.[22] Likewise, a patent and copyright search policy may inoculate an institution against the potential for copyright or patent infringement damages.[23]

In approaching this issue from a commons governance perspective, the negative externalities flowing from patent and copyright infringement could lead to uncertainty and eventual project abandonment.[24] As part of inductive coding, those interviewees that did describe rules and norms surrounding patent searching mentioned how one of their options was to abandon a project if they identified broad software patents that were being aggressively enforced. Accordingly, the failure to articulate a freedom to operate strategy could compromise the ability to identity what bioinformatics software can and cannot be reused in developing further derivative software. This blocking effect may hold true even where formal exceptions exist to permit experimental use with patented software.[25]

The challenge in integrating a freedom to operate searching policy into open source software governance remains the cost of engaging in freedom to operate. As Chapter Five demonstrates, there are inherent difficulties in identifying software and business method inventions from the patent literature.[26] However, both European and US interviewees described relying on and crucially forming collaborations with other researchers working in the same field through patent applications. Even though these were nascent relationships, they still demonstrate how patent disclosure can be used to establish research collaborations and exchange know-how. Accordingly, copyright notices and standardised Git tags might be used

---

(Working Paper No 24449, National Bureau of Economic Research, 2017) 34 <http://www.nber.org/papers/w24449>.

[21] Ross Anderson and Tyler Moore, 'Information Security: Where Computer Science, Economics and Psychology Meet' (2009) 367(1898) *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 2717–2727 2719.

[22] Ross Anderson et al., 'Measuring the Cost of Cybercrime' in *The Economics of Information Security and Privacy* (Springer, Berlin, Heidelberg 2013)265 268.

[23] Arti K. Rai, John R. Allison and Bhaven N. Sampat, 'University Software Ownership and Litigation: A First Examination' (2009) 87(5) *North Carolina Law Review* 1519–1570 1552.

[24] Jim David Flowers, *Explaining policy differences as a function of diverse governance institutions* (PhD Thesis, Georgia Tech, 2016) 271 <https://smartech.gatech.edu/handle/1853/54971>.

[25] See Kenneth Oye and Rachel Wellhausen's comments on the lack of a patent search strategy for synthetic biology. (Kenneth A. Oye and Rachel Wellhausen, 'The Intellectual Commons and Property in Synthetic Biology' in *Synthetic Biology* (Springer, Dordrecht 2009)121–140 126-7)

[26] James Bessen and Michael J. Meurer, *Patent Failure: How Judges, Bureaucrats, and Lawyers Put Innovators at Risk* (Princeton University Press, 2009) 198-99.

to indicate ownership.[27] Further, as discussed in Section 7.4 below, information rules and norms can be used in concert with a prior user rights defence to create defensive disclosures. This defensive approach may be particularly useful against overly broad patent claims on bioinformatics techniques that are already in widespread use. These approaches may provide bioinformaticians with a means of sharing know how about genuinely patentable inventions, whilst defending against aggressive patent claims.

> Recommendation One: Open source bioinformatics developers should use freely available patent search tools to identify bioinformatics patents. Although as a general rule patents are relatively rare in bioinformatics research, basic patent searches should be used when developing software with potential commercial value. In addition, academic or private sector participants in an open source bioinformatics project should disclose any associated patent applications, as well as the overlap between these applications.

7.3.2 *Improving Aggregated Decision Making Around Copyright and Patenting Enforcement for Bioinformatics*

The second category of institutional statements that were consistently not identified by interviewees across jurisdictions relate to aggregation rules. Aggregation rules pertain to how a single participant or multiple participants in a commons reach a decision on the management of a particular resource.[28] The inductive coding used in Section 6.3 of Chapter Six described how many group members of a bioinformatics software project made individual decisions on which licence to use. These institutional statements were in almost all cases explicitly articulated as norm or shared strategies. This 'open textualism', where the licence defines a set of foundational principles to govern ongoing interactions between the contracting parties,[29] can be contrasted with aggregation institutional statements. By contrast, aggregation institutional statements define how organisations collectively decide how a project should be licensed.

Even where there is a valid copyright claim to a project, each developer to the project must have the authority to enforce that copyright against infringers.[30] Some interviewees described how their academic employment contracts would mandate that all software of a certain

---

[27] Armijn Hemel and Shane Martin Coughlan, 'Making Sense Of Git In A Legal Context' (2018) 9(1) *International Free and Open Source Software Law Review* 19–33 32.

[28] As discussed in Section 6.3.1.5 of Chapter Six.

[29] Chen Wei Zhu, "Copyleft' Reconsidered Why Software Licensing Jurisprudence Needs Insights from Relational Contract Theory' (2013) 22(3) *Social & Legal Studies* 289–308 292.

[30] Andres Guadamuz and Andrew Rens, 'Comparative Analysis of Copyright Assignment and Licence Formalities for Open Source Contributor Agreements' (2013) 10(2) *SCRIPTed: A Journal of Law, Technology and Society* 207–230 214.

complexity developed during employment be assigned to the university. However, once copyright has been assigned to an employer, the assignee then has the rights to enforce that copyright. The assignee can then use that copyright in ways that the original developer may disagree with.[31] To this end, as described by interviewees, developers may nevertheless feel uncomfortable about assigning copyright to a third party for fear of creating a 'copyright troll'. By raising the barriers to reuse of software, this copyright assignment will in turn have an impact on downstream reuse and citation of software.[32] This concern was reflected by interviewees who expressed caution about assigning copyright to a third party as a means of funding the ongoing open source software development.

Likewise, in the context of patent acquisition, three interviewees expressed frustration at how patents on which they were listed as an inventor were drafted too broadly by patent attorneys or were commercialised in ways that they disagreed with. These results confirm both the perspective of many interviewees in this study as well as the findings of Zhen Lei, Rakhi Juneja and Brian Wright. Their study suggests that the majority of problems that flow from patents and copyright in a research context flow from the *enforcement* of these rights (as well as reach through rights implemented through agreements such as Material Transfer Agreements (MTAs)). By comparison, the mere *presence* of these rights was not reported as a problem by interviewees.[33] In the context of bioinformatics, version control software such as GitHub and the rapid pace of software development in general make it relatively straightforward for developers to bypass institutional technology transfer offices.

An additional complicating factor is policies surrounding ownership of inventions developed by a university's employee which may necessitate *jurisdictional* variation in aggregation rules. In the US, the *Bayh-Dole Act* was presumed to vest ownership of university developed inventions in the university employing the researcher.[34] However, the Supreme Court assessed university ownership of patents in *Stanford University v Roche Molecular Systems Inc*.[35] The Supreme Court held that the *Bayh-Dole Act* does not override the presumption under US patent law that the rights to an invention vest in the listed inventor in the absence of an explicit agreement to the contrary. In a similar fashion, in Australia Justice French of the Federal Court of Australia struck down an equivalent contract in *University of*

---

[31] Alberto Galasso, Mark Schankerman and Carlos J. Serrano, 'Trading and Enforcing Patent Rights' (2013) 44(2) *The RAND Journal of Economics* 275–312 276.

[32] James Howison and James D. Herbsleb, 'Scientific Software Production: Incentives and Collaboration' (Paper presented at *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* , 2011) 520.

[33] Zhen Lei, Rakhi Juneja and Brian D. Wright, 'Patents versus Patenting: Implications of Intellectual Property Protection for Biological Research' (2009) 27(1) *Nature Biotechnology* 36–40 39.

[34] Shubha Ghosh, 'Are Universities Special' (2016) 49(3) *Akron Law Review* 671–694 684.

[35] *Stanford University v Roche Molecular Systems, Inc* 131 S. Ct. 2188 (2011), 2192.

Western Australia v Gray.[36] In particular, the Federal Court held that a duty to conduct research is not sufficient to imply a term into research contracts that employee inventions are by default assigned to the university in question. Although professor's privilege was common amongst European countries until recently (and remains in Italy, Sweden and some UK universities), most European countries have transitioned to university ownership models.[37] Accordingly, any aggregation rules for open source projects must account for these institutional and jurisdictional differences.

One potential solution identified by US and EU interviewees was the assignment of rights to a third party trust. In an academic context, this trust is separate to the institution employing the researchers, and is established purely for the purpose of running the project. This institution will then be empowered to enforce that copyright on the contributor's behalf in the case of infringement. This 'trustee' approach is equivalent to foundations that have been created to support a number of different open source projects, including the Mozilla and Apache Foundations.[38] Equivalent projects have been established in recent times to try and create standardised file formats and software development for bioinformatics research. These institutes include the Open Bioinformatics Foundation or the Global Alliance for Genomic Health (GA4GH).[39] Although this strategy was not identified by Australian and New Zealand interviewees, it could drive open source development in Australia. However, in the absence of clear funding to support ongoing open source development, these institutional strategies may struggle to succeed. The next section discusses how funding strategies may affect the governance of different computational biology projects.

> Recommendation Two: Academic grant applications should be drafted so that the ownership of any copyrighted works or patents developed as part of the project are explicitly defined. If the development team wish to release open source software as part of their project, any associated licences should be listed within the grant application. The grant application should also include a notice of whether the development team intend to commercialise their software.

---

[36] *University of Western Australia v Gray (No 20)* [2008] FCA 498; *University of Western Australia v Gray* [2009] FCAFC 116.

[37] David B. Audretsch and Devrim Göktepe-Hultén, 'University Patenting in Europe: Does Faculty Ownership of Intellectual Property Impede University Technology Transfer?' in Albert N. Link, Donald S. Siegel and Mike Wright (eds.), *The Chicago Handbook of University Technology Transfer and Academic Entrepreneurship* (University of Chicago Press 2015) 191-5.

[38] Siobhán O'Mahony, 'Guarding the Commons: How Community Managed Software Projects Protect their Work' (2003) 32(7) *Research Policy* 1179–1198 1198-201.

[39] Charles M. Schweik and Robert C. English, *Internet Success: A Study of Open-Source Software Commons* (MIT Press, 2012) 33; Lillian L. Siu et al., 'Facilitating a Culture of Responsible and Effective Sharing of Cancer Genome Data' (2016) 22(5) *Nature medicine* 464–471 474.

7.3.3 *Adopting Novel Technology Transfer and Funding Strategies for Computational Biology Research*

As alluded to in the previous section, universities and funding agencies are increasingly emphasising the importance of uniform technology transfer standards for university research. The economic justification for the introduction of formalised technology transfer offices into universities is to provide a vehicle for driving applied research. This model prevents academics from having to work in the marketing of technology, or divesting themselves from conducting further research.[40] Within this model, the licensing fees from applied research can then be used to fund basic research.[41] As Tania Bubela and Timothy Caufield comment, this model is particularly effective at driving commercial biotechnology research due to the high rate of failure within biotechnology research.[42] This approach is confirmed by Pierre Azoulay's study into a random sample of academic patenting on biotechnology, which suggested that patenting had a positive impact on the rate of further publications.[43]

However, not all scientific fields are uniformly amenable to the patenting and technology transfer model observed in mainstream biotechnology research. Of the interviewees sampled for this thesis, only six described patents as a valuable mechanism for protecting a particular bioinformatics invention or signalling its value to inventors. These results suggest that patents are of limited utility as a mechanism for technology transfer. Further, as discussed in Section 5.3.3 of Chapter Five, the relatively small number of academic bioinformatics patents indicates the majority of filed bioinformatics algorithms do not meet the criteria for patent eligibility.[44] Finally, as discussed by interviewees who had filed for patents, patents are jurisdictionally bound. To achieve optimal coverage, patent applications must be filed for across a number of jurisdictions. This strategy is consistent with conventional logic on how patents should be used in portfolios.[45] However, the downside of this approach is that it is more expensive and risky than filing within single jurisdictions.

---

[40] Peter Van Dongen et al., 'The Relationships Between University IP Regimes, Scientists' Motivations and Their Engagement with Research Commercialisation in Europe' (2017) 8(2) *European Journal of Law and Technology* 2.

[41] Anthony D. So et al., 'Is Bayh-Dole Good for Developing Countries? Lessons from the US Experience' (2008) 6(10) *PLOS Biol* e262 2079.

[42] Tania M. Bubela and Timothy Caulfield, 'Role and Reality: Technology Transfer at Canadian Universities' (2010) 28(9) *Trends in Biotechnology* 447–451 447.

[43] Pierre Azoulay, Waverly Ding and Toby Stuart, 'The Impact of Academic Patenting on the Rate, Quality and Direction of (Public) Research Output' (2009) 57(4) *The Journal of Industrial Economics* 637–676 670.

[44] Either the standards for patentable subject matter as discussed in Section 3.3 of Chapter Three or the other requirements for patent eligibility, namely novelty, inventiveness, utility and adequate disclosure.

[45] See (Gideon Parchomovsky and R. Polk Wagner, 'Patent Portfolios' (2005) 154(1) *University of Pennsylvania Law Review* 1–77 1)

Nevertheless, including researchers in the technology transfer process may bypass some of the problems described above. One interviewee from Australia described a successful technology transfer process where the development team themselves played a significant role in designing the licensing model that was used for software.[46] An interviewee in the European Union also described the benefits for collaboration flowing from 'professor's privilege' (where rights vest with the inventor rather than the university).[47] In light of the analysis of the interviews in Chapter Six, this thesis recommends that ownership of software related inventions should vest with the developers. This ownership model would allow these inventors to pursue user driven technology transfer strategies that suit their specific research agendas. User inventors models can be contrasted with how some interviewees described themselves and colleagues being pressured into restrictive employment contracts that would restrict technology transfer.

In particular, from the analysis of interviews described in in Section 6.3 of Chapter Six there were two key technology transfer strategies beyond traditional patent protection that were used by bioinformaticians to encourage return on investment. The first possible model is a tiered dual licensing strategy, where academic users are provided with free access to the software in question but proprietary users are required to pay a nominated fee for use. This strategy has been used for both open source projects and for scientific tool licensing.[48] The effectiveness of a dual licensing model is supported by the patent publication pairs analysis in Section 5.3.3 of Chapter Five. In this analysis, even where a patent had been granted, the presence of that patent is not a prohibition on ongoing use of the initial open source project. However, the analysis of the interviews in Chapter Six demonstrated that developers who rely on dual licensing have clearly defined institutional rules as to when dual licensing is appropriate. This reluctance to rely on dual licensing is consistent with other research, which has revealed that dual licence enforcement can sometimes discourage software reuse.[49]

The second technology transfer strategy involves software as a service. A software as a service model would depend on bioinformaticians providing sequencing services using high

---

[46] This approach can be likened to Joel West's findings from a case study of Claude Shannon's theory of communications developed at MIT, noting researchers were more likely to ensure that this research was not only applied in an industrial context but also diffused through working papers and journal articles. (Joel West, 'Commercializing Open Science: Deep Space Communications as the Lead Market for Shannon Theory, 1960–73' (2008) 45(8) *Journal of Management Studies* 1506–1532 1524-5)

[47] This observation was supported by findings that 'professor's privilege' leads to greater academic and industry collaboration (Birgitte Andersen and Federica Rossi, 'Inefficiencies in markets for intellectual property rights: experiences of academic and public research institutions' (2012) 30(1) *Prometheus* 5–27 7).

[48] Peter Eckersley et al., 'Neuroscience Data and Tool Sharing' (2003) 1(2) *Neuroinformatics* 149–165 153-4; Tom Dedeurwaerdere, 'Databases, Biological Information and Collective Action' (Paper presented at *Social Informatics: An Information Society for all? In Remembrance of Rob Kling* , 2006) 166.

[49] Ville Oksanen and Mikko Välimäki, 'Free Software and Copyright Enforcement: A Tool for Global Copyright Policy?' (2006) 18(4) *Knowledge, Technology & Policy* 101–112 111.

performance computing hardware to support analysis. The software itself could then be released under an open source licence.[50] This strategy bypasses the need to handle the volume of transactions and advertising for ongoing proprietary licensing of software, which many interviewees admitted that their technology transfer offices were not prepared for. A variation on this licensing model is the development of 'bespoke' software for a particular researcher to solve a particular research problem. Bespoke software can be designed for an internal or external research team who might not otherwise have the necessary software engineering expertise.[51] These collaborations would be largely handled on a party to party basis between the developers and the external research team. Unfortunately, as two Australian interviewees mentioned, technology transfer offices and private research companies may view bespoke software development as a means of avoiding compliance with open source licences. The lack of contribution back into the main open source project then has the potential to undermine the sustainability of that project and create a fragmentation of standards.[52] In addition, the negative implications flowing from fragmentation demonstrate some of the limitations of private ordering strategies, as discussed in further detail below.

Recommendation Three: Academic bioinformatics developers, if commercialising publicly funded bioinformatics software, should prioritise non-proprietary technology transfer models. These technology transfer models can include the development of bespoke software and software as a service. Ultimately all academic commercialisation should be designed to support further development (as opposed to seeking a profit).

7.3.4 *Limitations of Private Ordering and Funding Strategies*

Despite the potential for private ordering strategies to improve collaboration in bioinformatics research, the interviews nevertheless revealed some limitations with respect to how copyright and patent law protects bioinformatics software. A particular example identified by three US interviewees is the applicability of copyright to software programs such as Docker and Juypter Notebook. These programs are designed to create a container, or an 'operating system within an operating system'. This operating system acts as an easy way of transferring scientific code and data, including custom bioinformatics workflows. It is possible to have many such docker containers for every dataset.[53] However, at least two US

[50] Vivek Navale and Philip E. Bourne, 'Cloud Computing Applications for Biomedical Science: A Perspective' (2018) 14(6) *PLOS Computational Biology* e1006144 2-3.

[51] James Howison et al., 'Understanding the Scientific Software Ecosystem and its Impact: Current and Future Measures' (2015) 24(4) *Research Evaluation* 454–470 459.

[52] Primavera De Filippi, 'Law of the Cloud: On the Supremacy of the User Interface over Copyright Law' (2013) 2(3) *Internet Policy Review* 5.

[53] Jack S. Hale et al., 'Containers for Portable, Productive, and Performant Scientific Computing' (2017) 19(6) *Computing in Science Engineering* 40–50 42.

interviewees expressed concern about the scope of copyright protection for the contents of Docker containers. These findings are consistent with US and Australian law on copyright not extending to compilations of data. For these interviewees, this question was particularly pertinent as the software and scripts that were included with these containers were often custom designed for that software. From these two examples, it is apparent that the boundaries of patent and copyright law will continue to stymie collaborative efforts. Accordingly, further legislative reform may be necessary to maximise the benefit that copyright law and patent law provide to commons based research strategies.[54] The next section will describe these reforms in further detail.

7.4 RECOMMENDATIONS FOR LEGISLATIVE AND ADMINISTRATIVE REFORM IN AUSTRALIA

7.4.1 *The Proposed Fair Use Defence and the Copyright Act 1968*

The first statutory recommendation flowing from this thesis is to introduce a fair use defence into the Australian *Copyright Act 1968* (Cth). Although a portion of the semi-structured interview questions related to patenting, interviewees were also questioned on their perspectives about the extent of copyright law. As Chapter Two discussed, copyright law remains both the primary form of protection for software, and the primary mechanism for enforcing open source licences. The decrease in the viability of software patents (at least in the US) following the decisions in *Bilski*,[55] *Mayo*[56] and *Alice*[57] may discourage patenting. Instead, software developers may rely on copyright as a more strategically appealing form of protection for software (and by extension bioinformatics algorithms).[58] For proprietary software developers, this shift has been reinforced by the recent decision of the Federal Circuit Court of Appeals in *Oracle v Google*.[59] This case has reversed thirty years of US precedent opposing the functional protection of software and also significantly narrowed the scope of the fair use defence for reverse engineering of computer software.[60] What are the implications of broadening the scope of copyright protection for software and narrowing the scope of the fair

---

[54] Eli Dourado and Alex Tabarrok, 'Public Choice Perspectives on Intellectual Property' (2015) 163(1) *Public Choice* 129–151 145.

[55] *Bilski v Kappos* 130 S. Ct. 3218 (2010).

[56] *Mayo Collective Services v Prometheus Laboratories* 132 S. Ct. 1289 (2012).

[57] *Alice Corporation v CLS Bank* 573 U.S. ¸134 2347 (2014).

[58] Christi J. Guerrini et al., 'Constraints on Gene Patent Protection Fuel Secrecy Concerns: a Qualitative Study' (2017) 4(3) *Journal of Law and the Biosciences* 542–564 555.

[59] *Oracle America, Inc. v Google Inc.* 750 F.3d 1339 (Fed. Cir. 2014).

[60] Peter S. Menell, 'API Copyrightability Bleak House: Unraveling and Repairing the Oracle v. Google Jurisdictional Mess' (2016) 31(3) *Berkeley Technology Law Journal* 1515–1596 1564.

use exemption? According to Wendy Gordon's theoretical economic analysis of an optimal copyright system, the amount of protection afforded to the copyrighted subject matter should be incomplete. Incomplete protection allows for the follow on creation of copyrighted works without requiring payment of licensing fees.[61] William Landes and Richard Posner argue that the fair use defence should be constrained to circumstances where the benefits of use exceed the costs of obtaining a copyright licence for the use of the work. In turn, this conclusion raises the question of what constitutes the optimal scope of fair use.[62]

Gordon's analysis of the *Universal Studios v Sony Corp (Betamax)*[63] copyright litigation, which concerned the use of the Betamax video recording software for private copying of movies, proves instructive. Gordon argues that opening a fair use defence to noncreative, exact copying would have the capacity to bypass market failure effects where socially beneficial licensing is impossible.[64] In more recent US copyright jurisprudence, this market failure doctrine can be traced to the importance of transformative use and partial copying, which are described by Sag as statistically significant indicators of a finding in favour of fair use.[65] A broad fair use exception that is not predicated on a finding of non-commercial use is crucial to allow the legitimate reuse of copyrighted works. This broad fair use exemption is particularly pertinent to the development of interoperable software, and particularly for bioinformatics software.[66] As Section 1.2 of Chapter One described, the trend in software development has been characterised by a shift away from a monolithic, sequential approach to software development towards an object oriented approach. Further, as Clark Asay notes, a broad fair use doctrine with respect to software development acts as a counter balance against the multiple copyright licences that may vest in any one software subsystem.[67]

The boundary and choice rules identified in this thesis demonstrate the strategies that software developers working in bioinformatics research use to make the determination on whether or not to incorporate particular pieces of software from outside their organisation. A major consideration identified through inductive coding in this study was the need to seek out

---

[61] Wendy J. Gordon, 'The Fair Use Doctrine: Markets, Market Failure and Rights of Use' in Richard Watt (ed.), *Handbook on the Economics of Copyright: A Guide for Students and Teachers* (Edward Elgar Publishing 2014) 91.

[62] William M. Landes and Richard A. Posner, 'An Economic Analysis of Copyright Law' (1989) 18(2) *Journal of Legal Studies* 325–364 358.

[63] *University City Studios, Inc. v Sony Corp. of America (Betamax)* 659 F.2d 963 (9th Cir. 1981).

[64] Wendy J. Gordon, 'Fair Use as Market Failure: A Structural and Economic Analysis of the "Betamax" Case and Its Predecessors' (1982) 82(8) *Columbia Law Review* 1600–1657 1601-2.

[65] Matthew Sag, 'Predicting Fair Use' (2012) 73(1) *Ohio State Law Journal* 47–92 75.

[66] Vincent J. Carey and Victoria Stodden, 'Reproducible Research Concepts and Tools for Cancer Bioinformatics' in Michael F. Ochs, John T. Casagrande and Ramana V. Davuluri (eds.), *Biomedical Informatics for Cancer Research* (Springer US 2010)149–175 168.

[67] Clark D. Asay, 'Software's Copyright Anticommons' (2016) 66(2) *Emory Law Journal* 265–332 270.

certain useful statistical functions or certain ontologies for representing scientific workflows. Although the need to ensure homogenous or at least consistent licensing for all components in a software package was an important consideration (as evidenced by the rules and norms that were identified by participants), developers still noted that certain packages were more important than others (such as support vector machines algorithms) and were therefore more difficult to reverse engineer.[68] Interviewees described a number of strategies to bypass incompatible licensing, such as avoiding implementing the package, reimplementing the software or requesting relicensing. Nevertheless, that functionality may nevertheless be unavailable where copyright protection for software is interpreted too broadly, such as in the case of *Oracle v Google*.

Many open source licences are written with explicit reference to the fact that they do not deny the licensee their rights under the fair use doctrine.[69] For example, there are variations on certain open source licences, such as the GPL Affero licence. The GPL Affero licence allows a licensee to use a GPL licensed software through APIs in a web service without activating the copyleft provisions of the GPL. The Affero licence is therefore designed on the premise that the copyright licence in the standard GPL extends to the use of GPL licensed libraries in larger software compilations.[70] Nevertheless, outside of explicit relicensing under open source licences and where the licensee cannot give permission to reuse copyrighted materials, the decision in *Oracle v Google* could act as a bar on the reverse engineering of functionality by preventing other developers from engaging in non-literal reverse engineering.[71] In particular, this precedent could bar the reimplementation of functionality of proprietary licensed software, as well as decrease the incentive of the licence holders of that software to cooperate with open source developers.

Unfortunately, limitations on the reverse engineering of software also exist outside of the US. As part of the 2016 inquiry into intellectual property laws in Australia, the Productivity Commission of Australia recommended introducing a fair use exemption into the *Copyright Act 1968* (Cth). Likewise, although there is no tabled legislation or impetus for reform at present, there is current debate in academia in New Zealand about whether a fair use defence should be introduced into the *Copyright Act 1994* (NZ).[72] The fair dealing exemptions

---

[68] These findings are consistent with the idea that open source software is frequently developed to meet a particular developer need (Audris Mockus, Roy T. Fielding and James D. Herbsleb, 'Two Case Studies of Open Source Software Development: Apache and Mozilla' (2002) 11(3) *ACM Trans. Softw. Eng. Methodol.* 309–346 325).

[69] Lothar Determann, 'Software Copyright's Oracle from the Cloud' (2015) 30(1) *Berkeley Technology Law Journal* 161 175.

[70] Primavera De Filippi and Miguel Said Vieira, 'The Commodification of Information Commons: The Case of Cloud Computing' (2014) 16(1) *Columbia Science and Technology Law Review* 102–143 135.

[71] Henry Chesbrough and Marshall Van Alstyne, 'Permissionless Innovation' (2015) 58(8) *Communications of the ACM* 24–26 25.

[72] Alexandra Sims, 'The Case for Fair Use in New Zealand' (2016) 24(2) *International Journal of Law and*

contained in Division 4A of the *Copyright Act 1968* (Cth) provide for limited exceptions to copyright infringement in computer software for normal use or study, producing interoperable programs or correcting errors.[73] However, as Cristina Cifuentes and Anne Fitzgerald note (and as subsequently confirmed in *CA Inc v ISI Pty Ltd*[74]) the interoperable software development exemption is very narrow in scope, and only applies where the licensee has obtained a valid licence from their licensor.[75] This requirement acts as a significant bar on the reimplementation of software functionality, which is a key aspect of open source development.

One potential solution to this problem is to include specific statutory recognition of open source licences. [76] However, as Nicolas Binctin notes, these reforms simply confirm that open source licensing is possible and are therefore largely symbolic.[77] Within a common law framework a stronger legislative reform to rectify the lack of flexibility is to introduce a norm based fair use exception. This recommendation is supported by the interview results in Chapter Six, where interviewees frequently mentioned that they were reliant on the good will of other developers to relicense software. In addition, this recommendation is supported by Christina Handke, Lucie Guibault and Joan-Josep Vallbé's survey of the impact of copyright exceptions for academic data mining research in the EU. This survey suggests that copyright regimes requiring express consent from all rights holders leads to significantly less research outputs from data mining research.[78] Further, with the introduction of narrow exceptions for text and data mining under the revised Copyright Directive, it is unlikely a broader test for fair use will emerge under European Union law. In particular, the chilling effects of a restrictive interpretation of fair use exceptions are reflected in Thomas Cotter's 'rights accretion' hypothesis. This hypothesis states a legitimate user of copyrighted material could be discouraged from using that material due to concerns about the breadth of the fair use defence. Accordingly, neither party benefits due to the cost of cross licensing and the unavailability of

---

Information Technology 176–202 178.

[73] *Copyright Act 1968* (Cth) ss47B, 47D, 47E.

[74] *CA Inc v ISI Pty Ltd* (2012) 201 FCR 23.

[75] Anne Fitzgerald and Christina Cifuentes, 'Accommodating Computer Software to Copyright Doctrine: Defining the Scope of Copyright Protection for Software' (2000) 11(2) *Journal of Law and Information Science* 224–253 236; Australian Law Reform Commission, *Copyright and the Digital Economy*, Final Report No 122 (November 2013) 370.

[76] Such as the legislative reforms introduced into the French Copyright Act (*Code de la propriété intellectuelle 1992* (French Intellectual Property Code) Article 122-7-1)

[77] Nicolas Binctin, 'The French Copyright Law Opens Its Arms to the FOSS' in Axel Metzger (ed.), *Free and Open Source Software (FOSS) and other Alternative License Models* (Springer, Cham 2016)477–493 204.

[78] Christian Handke, Lucie Guibault and Joan-Josep Vallbé, 'Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research' in *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science: Scale, Openness and Trust: proceedings of the 19th International Conference on Electronic Publishing* (2015)120 127-9.

fair use as a defence.[79] Whilst some interviewees mentioned that they were able to persuade other developers to relicense their source code as part of a larger project, in part the effectiveness of these requests rested with the fact that functionality was relatively easy to reimplement.

If the current fair dealing exceptions in the *Copyright Act 1968* (Cth) are to be replaced with a US style fair use exception, it must be constructed to support a broad approach to fair use with respect to reverse engineering of functionality. How might this broad approach be achieved? The limited number of jurisdictions that have implemented a fair use defence into their national copyright regimes have largely relied on US precedent.[80] However, such an approach risks importing the problematic precedent of *Oracle v Google* into Australia. In addition, the US Trade Representative is currently opposed to the diffusion of fair use on an international level.[81] Finally, as discussed in Section 2.3.3 of Chapter Two, there may be inconsistency between the fair use exception and the three part test under Article 13 of the TRIPS Agreement. This incompatibility may arise from the uncertainty and breadth of the defence, as well as the fact that it nullifies legitimate trading rights.[82] This question of invalidity is particularly pertinent in light of the fact that more recent US case law on fair use has permitted uses that may affect the plaintiff's market share or entry into a market. The potential of this effect would therefore create further conflict with the first part of the Article 13 test.

Where to then for the introduction of a fair use defence into Australian copyright law? UK legislation might provide an alternative model towards a fair use defence. The *Copyright, Patents and Designs Act 1988* (UK) is akin to the current *Copyright Act 1968* (Cth) in that it contains explicit fair dealing exceptions for private research and non-commercial use, criticism or review of a work and reporting of current affairs and news.[83] However, there is conflict as to whether these sections should be interpreted broadly or narrowly, with Lord Denning in the case of *Hubbard v Vospar* comparing the application of the fair dealing exemptions to the equitable nature of fair comment under defamation law.[84] The more recent judgment of Justice Proudman in *Newspaper Licensing Agency v Meltwater BV Ors*[85] adopted a narrower interpretation of

---

[79] Thomas F. Cotter, 'Fair Use and Copyright Overenforcement' (2007) 93(4) *Iowa Law Review* 1271–1318 1277-8.

[80] Gabriel J. Michael, *"To Promote the Progress"? Explaining the Global Diffusion of Intellectual Property Law* (PhD Thesis, The George Washington University, 2014) 238-52.

[81] Jonathan Band and Masanobu Katoh, *Interfaces on Trial 2.0* (MIT Press, 2011) 182.

[82] Ruth L. Okediji, 'Toward an International Fair Use Doctrine' (2000) 39(1) *Columbia Journal of Transnational Law* 75–176 126-34.

[83] *Copyright, Designs and Patents Act 1988* (UK) sections 29-30.

[84] *Hubbard v Vosper* [1972] 2 QB 84, 94.

[85] In this decision, Justice Proudman noted that the broad interpretation of fair dealing promulgated by Lord Denning was contradicted by the UK's requirements under the InfoSoc Directive. (*Newspaper Licensing Agency v Meltwater BV & Ors* [2010] EWHC 3099, paragraph 119)

the meaning of what was fair for the purposes of fair dealing than Lord Denning in *Hubbard v Vospar*.

Nevertheless, Antony Dnes argues that it may be possible to combine the fair dealing exemptions with other specific exceptions within the *Copyright, Patents and Designs Act 1988* (UK) to create a 'fair use' jurisprudence.[86] In addition to the fair dealing exemptions contained in sections 29 to 30, section 31 of the UK codifies a *de minimis* exception for the taking of insubstantial pieces of copyrighted material. Equivalent provisions exist in other common law copyright doctrines, including Australian and New Zealand copyright law.[87] Dnes argues that in concert with the broad interpretation of fair dealing supported by Lord Denning in *Hubbard*, the *de minimis* exception could be used to build the foundations of a fair use jurisprudence in the UK.[88] This style of fair use permits a flexible approach when dealing with new technologies such as generic text mining software as well as dedicated bioinformatics software which may otherwise be compromised by restrictive copyright standards. Aside from the interviews presented in this thesis, this finding is demonstrated by the findings of Handke, Guibault and Vallbé. They note that publications that cite data mining technology make up a higher proportion of total scientific publications from EU member states that do not require express consent to reuse copyrighted material than those that do.[89]

Further, as Kylie Pappalardo and Brian Fitzgerald note, UK copyright law has significantly influenced the development of Australian copyright law.[90] Indeed, in determining the scope of this section of the Commonwealth Constitution, the High Court has repeatedly characterised these powers as embodying a social contract model of copyright (and other forms of intellectual property right), where the interests of the rights holder must be balanced against the interests of the public.[91] From this flexible approach, it may be possible to implant a fair use doctrine into Australian copyright law that does not unfairly prejudice the legitimate interests of software developers in writing interoperable software.

---

[86] Antony W. Dnes, 'Should the UK Move to a Fair-Use Copyright Exception?' (2013) 44(4) *IIC - International Review of Intellectual Property and Competition Law* 418–444 426, 430.

[87] *Copyright Act 1968* (Cth) section 14.

[88] Antony W. Dnes, above n86, 444.

[89] Christian Handke, Lucie Guibault and Joan-Josep Vallbé, above n78 127-9.

[90] Particularly with respect to the scope of the Commonwealth's intellectual property powers under section 51(xviii) (Kylie Pappalardo and Brian Fitzgerald, 'Copyright, Fair Use and the Australian Constitution' in Brian Fitzgerald and John Gilchrist (eds.), *Copyright Perspectives: Past, Present and Prospect* (Springer 2015)125 129-30).

[91] *Grain Pool of Western Australia v Commonwealth of Australia* (2000) 202 CLR 479, 496; *IceTV Pty Ltd v Nine Network Australia Pty Ltd* (2009) 239 CLR 459, 472 (French CJ, Kiefel and Crennan JJ jointly held that the 'information/expression dichotomy, in copyright, is rooted in considerations of social utility').

Recommendation Four: Australia copyright law should move away from express fair dealing exceptions towards fair use based exemptions. In particular, the *de minimis* exception should be used as the legislative conduit by which fair use can be introduced into Australian copyright law. In addition, any reform to introduce a fair use defence into copyright law should include a non-exhaustive list of acts that constitute fair use. These acts should include the reimplementation of minimal software functionality to improve interoperability and the use of software for data mining and text retrieval.

However, as discussed throughout this thesis, copyright is not the only means of protecting software; it is also necessary to consider the impact of patent reform on bioinformatics software.

7.4.2 *Are Reforms to Patent Law Necessary to Protect Open Source Computational Biology Research in Australia?*

7.4.2.1 Patentable Subject Matter

The final two sections of this chapter address the question of the boundaries of patent protection and patent infringement for bioinformatics software. This section considers what legislative strategies can be used to reduce the negative impact of patents identified by interviewees. First turning to address the question of patentable subject matter, the European Union and New Zealand have decided to prohibit software patents 'as such'.[92] The Australian Productivity Commission also recommended in an earlier interim report placing 'software and business method' inventions outside the scope of patentable subject matter under the *Patents Act 1990* (Cth).[93] This decision was eventually reversed in the Commission's final report, due to concerns from patent stakeholders about the effect of an explicit exclusion on the technological neutrality of the patent system.

Nevertheless, the Australian Patent Office (APO) appears to have taken matters into its own hands and has adopted the 'technical effect' test from the UK Court of Appeal's decision in *Aerotel Holdings Ltd; Macrossan's Application* ('*Aerotel/Macrossan*').[94] The Australian Patent Office has justified this approach on the grounds that the Court of Appeal approached the question of patentability in a similar fashion to the High Court in *D'Arcy v Myriad*.[95] In particular, the Australian Patent Office has made heavy reference to the reasoning of the plurality of Chief Justice French, and Justices Kiefel, Bell and Crennan in *D'Arcy* . This plurality held that in determining whether a particular invention fell within the bounds of the

[92] As discussed in Section 3.3.3.1 of Chapter Three.

[93] Productivity Commission, *Intellectual Property Arrangements (Draft)*, Draft Report No 78 (2016) 249-52.

[94] [2006] EWCA Civ 1371. IP Australia, 'Manual of Practice and Procedure' (last update 1 August 2017), s. 2.9.2.2

[95] *D'Arcy v Myriad Genetics Inc* (2015) 258 CLR 334.

'manner of manufacture' test, reference needed to be made to the patent laws of other countries.[96]

Although the APO's rulings are not binding, it would appear that it is attempting to unify Australian rules on software patent examination with the four step test presented in *Aerotel/Macrossan*[97], and subsequently confirmed in *HTC v Apple*[98]. The effect of this test is that pure software and business method patents are excluded from the scope of patent eligibility except where novel and inventive contribution contains a technical effect. Phillip Leith cautions that patent attorneys will continue to use novel strategies to bypass statutory restrictions on patentable subject matter, such as filing under different classes and modifying patent claims.[99]

However, Brendon Beheshti argues that because the US decision in *Alice* only requires that the patent disclose a technical effect without pinpointing any part of the invention that needs to include this technical effect, whereas *Aerotel/Macrossan* requires the technical effect to relate to the original contribution to the literature, *Aerotel/Macrossan* is a more useful guide than *Alice* for determining when software will be patentable.[100] Further, the fact that patent offices are moving towards this heightened standard for patentable subject matter is indicative that many of the concerns around patent thickets are likely to decrease significantly. This finding is confirmed by the analysis of semi-structured interviews in Chapter Six, where only two academic bioinformaticians mentioned that they were involved in patent disputes. Accordingly, this thesis agrees with this unification and argues strongly for Australia to retain the four step test as the appropriate test for determining patent eligibility for computer implemented inventions. In particular, this test will likely have the desired effect of discouraging the proliferation of low quality software patents.

> Recommendation Five: Patent offices should attempt to unify the test for software patent eligibility with that established by the Court of Appeal of England and Wales in *Aerotel/Macrossan*. In particular, in determining the scope of patentable subject matter, there must be a technical effect that elevates the software claim beyond intangible subject matter. Further, this technical effect must constitute the inventive element of the software.

---

[96] *D'Arcy v Myriad Genetics Inc* (2015) 258 CLR 334, 19.

[97] *Aerotel Ltd v Telco Holdings Ltd Ors* [2006] EWCA Civ 1371, paragraph 40.

[98] *HTC Europe Co Ltd v Apple Inc* [2013] EWCA Civ 451, paragraphs 35-36.

[99] Philip Leith, 'Patenting Programs as Machines' (2007) 4(2) *SCRIPTed: A Journal of Law, Technology and Society* 214–226 220.

[100] Brendon Beheshti, 'Getting beyond Abstract Confusion: How the United Kingdom's Jurisprudence Can Aid in Developing an Analytic Framework for Patent-Eligibility in Light of Alice v. CLS Bank' (2014) 10(2) *Washington Journal of Law, Technology & Arts* 137–152 148-51.

7.4.2.2 Patent Exceptions and Defences

The variable scope of research exceptions in each jurisdictions presents unforeseen challenges for software development, particularly where this development is conducted on a global basis. As discussed in Section 3.3.3.2 of Chapter Three, a notable recent law reform in both Australia and New Zealand has been the introduction of an experimental use exception into patent law, which permits research *on* a patented invention.[101] However, research by Dianne Nicol and colleagues suggests that the Australian research exception in fact offers a more restrictive interpretation of norms which already exist in academic research.[102]

Further, in Australia there is ongoing uncertainty as to whether the use of patented software in scientific research amounts to use on the patented invention (which is permitted under Section 119C) or research with the patented invention (which is not permitted under Section 119C). As Chapter Three discussed, this question is particularly vexing given that software is often classified as an 'enabling' technology with multiple different potential applications. These applications may include both research and commercial purposes. In response to this conundrum, Julie Cohen and Mark Lemley suggest introducing an 'interoperability' exemption into patent law to protect legitimate reverse engineering of patented algorithms.[103] Support for this 'interoperability exemption' emerges from the interview results in Part 6.3.1.3 of Chapter Six. In particular, there was limited evidence of interviewees avoiding using patented software or being unable to obtain a licence for such software.

> Recommendation Six: The operation of experimental use defences (such as under section 119C of the *Patents Act 1990* (Cth)) should be clarified to explicitly permit the development of software that uses otherwise patented software for research purposes. For example, in the Australian and New Zealand *Patent Acts*, research with a patented invention should be defined so as to include but not be limited to the development of interoperable software for scientific research purposes.

An additional complement to the experimental use exception, as discussed in Chapter Three, would be reliance on a prior rights defence. Both US and Australian patent law currently feature a prior user rights defence.[104] In the US, the prior rights defence was first issued for business method patents only in light of the *State Street Bank v Signature Financial*

---

[101] *Patents Act 1990* (Cth) section 119C; *Patents Act 2013* (NZ) section 143.

[102] Dianne Nicol et al., *The Innovation Pool in Biotechnology: The Role of Patents in Facilitating Innovation* (University of Tasmania, 2014) 181.

[103] Julie E. Cohen and Mark A. Lemley, 'Patent Scope and Innovation in the Software Industry' (2001) 89(1) *California Law Review* 1–57 16-33.

[104] *35 U.S. Code §273 - Defense to infringement based on prior commercial use*; *Patents Act 1990* (Cth) section 119.

Group Inc[105] decision. The purpose of this defence was as a means of immunising potentially innocent infringers from patent litigation.[106] This defence was then broadened for all patents following the passage of the *America Invents Act*. The *America Invents Act* shifted the US from a 'first to invent' to a 'first to file' system for determining patent ownership, therefore harmonising the US patent system with most other jurisdictions.[107] Supporters of prior use rights such as Katherine Strandburg argue that it overcomes a key weaknesses of the patent system with respect to software; that is, determining whether use of a particular algorithm infringes on a patented algorithm.[108] Further, Greg Vetter, along with Shubha Ghosh, notes that the *America Invents Act* also broadened the scope of the 'commercial uses' definition under § 273 to include other uses, including uses by nonprofit entities such as universities and hospitals.[109] This definition would presumably be broad enough to cover the prior use of bioinformatics software in academic and hospital environments, as described in Section 7.3.2.

The EU and the rest of the world have been historically a 'first to file' patent systems, as opposed to the US, which has been historically a 'first to invent' patent system. Accordingly, prior user rights have been a feature of supranational IP rights outside of both jurisdictions.[110] In the UK, section 64 of the *Patent Act 1977* (UK) permits use or intentions to put into use of a patented invention prior to the priority date of the patent. The legality of this use is conditional provided that the use of that patented method being conducted in good faith.[111] Crucially this definition is not limited to industrial use of the invention. As the English Court of Appeals held in *Lubrizol v Esso Petroleum*,[112] there must be an infringing act or preparation to perform an infringing act that amounts to disclosure. The prior user rights exception under the *Patents Act 1990* (Cth) also permits a prior user to exploit the allegedly patented invention in a fashion that they did before the priority date of that invention.[113]

---

[105] *State Street Bank Trust Co. v Signature Financial Group Inc.* 149 F. 3d 1368 (Fed. Cir. 1998).

[106] Gaia Bernstein, 'The Rise of the End User in Patent Litigation' (2014) 55(5) *Boston College Law Review* 1443–1500 1448-9.

[107] *American Inventors Protection Act 1999* (Public Law 106-113); *America Invents Act 2011* (Pub L No 112-129) section 5.

[108] Katherine J. Strandburg, 'What if There Were a Business Method Use Exemption to Patent Infringement' (2008) *Michigan State Law Review* 245–278 256.

[109] *35 U.S. Code §273 - Defense to infringement based on prior commercial use* (c)(1) and (2); Greg R. Vetter, 'Are Prior User Rights Good for Software' (2014) 23(3) *Texas Intellectual Property Law Journal* 251–312 292; Shubha Ghosh, above n34, 690.

[110] *Convention on the Grant of European Patents*, opened for signature 5th October 1973, 1065 UNTS 254 (entered into force 7th October 1977) Article 2(2) in connection with Article 64(1); *Patentgesetz [German Patent Act] 1980* § 12 Abs. 1.

[111] *Patents Act 1977* (UK) section 64.

[112] *Lubrizol Corporation & Anor v Esso Petroleum Co Ltd* [1998] EWCA Civ 744, 761.

[113] *Patents Act 1990* (Cth) section 119; Jane Nielsen and Dianne Nicol, 'Whither Patent Use without Authorization in Australia' (2008) 36(3) *Federal Law Review* 333–364 336.

Whilst section 119 requires continuous use of the invention (and does not permit abandonment of the invention), there is case law to suggest that the disclosure of source code (as opposed to the demonstration of a particular software package) may be enough to establish prior use.[114] Accordingly, the prior use rights defence in Australia, in concert with the experimental use defence under section 119C, represents a mechanism by which academic bioinformatics developers could potentially immunise themselves from patent claims for existing open source software.

> Recommendation Seven: In the alternative, the operation of the prior use defence (such as under section 119 of the *Patents Act 1990* (Cth)) should be clarified. In particular, this defence should encompass the use of open source software that was in widespread use prior to the acquisition of a patent licence.

One problem though, as Martin Schmidt notes, is that the prior user rights defence is purely territorial and is limited to the jurisdiction where the prior use occurred.[115] In other words, a prior user rights defence will only provide an incomplete defence for bioinformatics researchers against the use of existing bioinformatics methods. Further, as discussed in Section 6.3.2.3 of Chapter Six, interviewees explained that for the most part, patents had limited utility for commercialising software. In particular, interviewees reported that the cost and time taken to acquire a patent was too great, given the rapid speed of bioinformatics development. This finding creates a peculiar situation where patents neither pose a significant risk to academic bioinformatics research nor act as an incentive to conduct research. This finding is also supported by the relatively small number of academic patents identified in Chapter Five, particularly in Australia. The final section of this chapter therefore examines whether a novel, *sui generis* regime could be suitable as a means of protecting software and other emerging technology. As discussed in Section 2.2.2 of Chapter Two, as well as Section 3.5.3 of Chapter Three, the idea of *sui generis* protection for software has been considered but rejected in favour of established intellectual property regimes. Nevertheless, the findings of this thesis suggest that a sui generis regime could be designed to provide protection for software functionality without the need to apply for a patent. This recommendation for reform is more certain than the other two recommendations given previous rejections of *sui generis* regimes. However, in Australia, this *sui generis* regime may be implemented as a replacement for the current innovation patent available which exists under the *Patents Act 1990* (Cth).

---

[114] *Jupiters Ltd v Neurizon Pty Ltd* [2005] FCAFC 90, 140.

[115] Martin P. Schmidt, 'Patent Strategies in the Process-Related Industries: Outline of the Problems' (2013) 43(3) *R&D Management* 242–251 243.

7.4.3 *Sui Generis Regimes and 'Partial Patents' - Is there Space for a Novel Intellectual Property Regime for Software and Bioinformatics?*

As both the doctrinal analysis from Chapters Two and Three discussed, neither copyright or patent law provides optimal protection for software programs. Whilst copyright was originally the *de facto* regime for the protection of software, it has traditionally only protected the literal text of software. By contrast, the functional elements of that software that frequently represent the 'innovative' part of software are not protected by copyright. The *Oracle v Google* decision, as well as the interviews in Chapter Six, demonstrate that there is still substantial uncertainty over what amounts to appropriate reuse of copyrighted software. This uncertainty holds true even where this software is released under an open source licence. By comparison, patent law ostensibly provides greater protection for the functional elements of software, such as user interfaces and algorithm design. However, as the analysis of the interviews in Chapter Six demonstrates, 'pure' software patents may undermine the development of interoperable software, as they are jurisdictionally bound, costly to acquire and difficult to enforce. For this reason, for the majority of interviewees across jurisdictions a primary justification for choosing to release software under an open source licence (instead of a proprietary licence) was the costs of acquiring a patent. Despite the important role that open source licences play in ameliorating the transaction costs associated with bioinformatics development, neither regime represents an ideal mechanism for protecting software.

One potential solution to the dilemma is the introduction of a *sui generis* regime for the protection of software. Prior to the introduction of copyright protection for software, both the EU and Japan were openly considering implementing *sui generis* software protection.[116] As early as the 1990s the idea of a *sui generis* regime for software received significant academic support from legal academics such as Pamela Samuelson and Jerome Reichman and computer science academics such as Randall Davis.[117] Although the *TRIPS Agreement* stymied these ambitions, subsequent research by John Swinson, Anton Hughes and Vikrant Vaseudeva has suggested the introduction of a *sui generis* regime into US, Australian and Indian law respectively. This regime would exist as an alternative to patent protection.[118] Simone Rose has also recommended a *sui generis* model for the protection of 'isolated bioproducts' as an

[116] As discussed in Section 2.2.2 of Chapter Two.

[117] Pamela Samuelson et al., 'A Manifesto concerning the Legal Protection of Computer Programs' (1994) 94(8) *Columbia Law Review* 2308–2431 2412-3.

[118] John Swinson, 'Copyright or Patent or Both: An Algorithmic Approach to Computer Software Protection' (1991) 5(1) *Harvard Journal of Law & Technology* 145–214 211; Anton Hughes, 'Avoiding the Software Patent Problem: An Alternative Fix for TRIPS Junkies' (2007) 14(1) *eLaw: Murdoch University Electronic Journal of Law* 100–116 115; Vikrant Narayan Vasudeva, 'A Relook at Sui Generis Software Protection Through the Prism of Multi—Licensing' (2013) 16(1-2) *The Journal of World Intellectual Property* 87–103 99.

alternative mechanism to patent protection in the aftermath of the *AMP v Myriad* decisions.[119] The objective of this 'partial patent' *sui generis* regime is to overcome the high transaction costs associated with acquiring formal intellectual property for open innovation. In addition, a *sui generis* partial patent regime could represent a ground up approach for modelling patent law around open source development practices.[120]

Similar concepts have been explored by other legal scholars in the US and in the EU.[121] However, Australia could be an ideal economic environment to test the economic feasibility of this model, particularly in light of the Productivity Commission's recommendation to entirely abolish the innovation patent system.[122] The Productivity Commission's primary criticism of the innovation patent system is that it is under utilised by its intended target audience (that is, innovative small and medium enterprises). Instead, low value innovation patents are being used strategically to discourage innovation due to the uncertainty over their scope.[123] The Productivity Commission's report examined the economic value of other utility models or second tier patent systems in other jurisdictions. In particular, the Commission focused on utility models which are regarded as contributing substantial economic, value such as the German utility model system. However, as the Productivity Commission concluded, other economic studies into these patent systems suggested that utility model patent systems only increase the speed of the diffusion of new technology. Further, these regimes do not encourage the production of innovative novel technology.[124] The equivocal evidence regarding the capacity of innovation patents to incentivise innovation was a key rationale behind the Productivity Commission's recommendation to abolish them.

There is also substantial inconsistency in the scope of subject matter that can be protected under even similar utility model systems. For example, algorithms are not protected subject matter under German utility model law, whereas under Austrian utility model law they are.[125]

---

[119] Simone A. Rose, 'Semiconductor Chips, Genes, and Stem Cells: New Wine for New Bottles?' (2012) 38(1) *American Journal of Law & Medicine* 113 154-5.

[120] This approach might also be justified on the grounds of the interviews that have been conducted for this study (Chen Wei Zhu, 'A regime of droit moral detached from software copyright?—the undeath of the 'author' in free and open source software licensing' (2014) 22(4) *International Journal of Law and Information Technology* 367–392 379-80).

[121] Gideon Parchomovsky and Michael Mattioli, 'Partial Patents' (2011) 111(2) *Columbia Law Review* 207; Geertrui Van Overwalle et al., 'Inventing Inclusive Patents. From Old to New Open Innovation' in *Kritika* (2015)1206–277.

[122] Productivity Commission, *Intellectual Property Arrangements*, InquiryReport No 78 (2016) Chapter 7, 239-62, recommendation 7.1.

[123] Benjamin Mitra-Kahn et al., 'The economic impact of innovation patents' (Economic Research Paper No 5, IP Australia, May 2015) 14, 24; above n122, Chapter 7, 253.

[124] Alfred Radauer et al., *Study on the Economic Impact of the Utility Model Legislation in Selected Member States: Final Report* (Publications Office, 2015) 37-42.

[125] Uma Suthersanen, *Utility Models and Innovation in Developing Countries* (International Centre for Trade and Sustainable Development (ICTSD), 2006) 14.

Accordingly, the justifications for retaining the innovation patent system in Australia are limited. The answer to this dilemma lies in the objective of providing a formal governance mechanism for open innovation through a *sui generis* intellectual property regime. This *sui generis regime* would offer a reduced term of protection (of approximately ten years) in exchange for lowered examination requirements. Although no interviewees in Europe or Australia mentioned engaging with the innovation patent system, there is precedent within plant breeders rights protection. In particular, Article 27.3(b) of the *TRIPS Agreement* permits nation signatories to create an 'effective *sui generis*' system. To quote Jack Kloppenburg, this Article allows national signatories with 'an opportunity to shape legislation to protect the interests and needs of farmers and indigenous peoples, and to craft [intellectual property right] arrangements that respect and reward collective invention'.[126] Further, as Peter Cummings notes, the German utility model system has undergone considerable reform since its inception to transform it from an economic complement rather than an alternative to patent protection.[127] This suggests that there is scope to use the legislative framework for the innovation patent system as a mechanism within which to embed a *sui generis* system of protection.

How then should this *sui generis* regime be designed so as to appropriately encourage innovation in software engineering? As Richard Gilbert and Carl Shapiro note, the question of optimal patent protection focuses on two factors: the breadth of the patent (that is, how much to reward the patentee) and the length of the patent (that is, how to structure each reward).[128] Firstly, exceptions to copyright and patent infringement are designed to legitimise use of a patented invention where the transactional costs of use prohibit socially beneficial licensing. In particular, the absence of information rules with regarding patent searching and aggregation rules regarding standards setting suggest a broader disengagement with the patent system by academic bioinformaticians. In part, this disengagement can be attributed to how interviewees viewed the value of software patents relative to physical patents such as patents on software embedded in computer hardware or genome editing systems.[129] In particular, interviewees noted that transaction costs of obtaining patent protection were so high as to discourage any form of socially beneficial licensing using patents. This is particularly pertinent, as conventional wisdom dictated that patents are best acquired as portfolios rather than single patents.

---

[126] Jack Kloppenburg, 'Impeding Dispossession, Enabling Repossession: Biological Open Source and the Recovery of Seed Sovereignty' (2010) 10(3) *Journal of Agrarian Change* 367–388 372.

[127] Peter A. Cummings, 'From Germany to Australia: Opportunity for a Second Tier Patent System in the United States' (2009) 18(2) *Michigan State University College of Law Journal of International Law* 297–322 304-5.

[128] Richard Gilbert and Carl Shapiro, 'Optimal Patent Length and Breadth' (1990) 21(1) *The RAND Journal of Economics* 106 106.

[129] As demonstrated by the quote in Section 6.3.1.3 of Chapter Six.

Secondly, an additional observation arising from the semi-structured interviews in Chapter Six is that academic bioinformaticians are reluctant to seek patents as a means of setting standards. When questioned, no interviewees discussed their patents being used to help set standards in bioinformatics research. Interviewees attributed this avoidance to the cost of obtaining and maintaining such patents, in addition to the strong prevailing norms in favour of open as opposed to patent oriented standards. The minimal standardisation approach described by interviewees can be contrasted with the way in which standards based patent licensing is frequently used in industrial computing and semiconductor research.[130] In their 2013 report, Keith Maskus and Steven Merrill make a particular point that patent pledging in the bioinformatics research community appears to be unusual. By contrast, most research occurs within academic communities as opposed to standard setting organisations. These findings were confirmed by the semi-structured interviews conducted for this research project, despite Maskus and Merrill suggesting a potential space for standard setting with respect to genetic data structures.[131]

The turbulence of the bioinformatics and computational biology development field undermines the effectiveness of patenting as a means of protection. This instability creates a flow on effect into determining the information and aggregation rules for patent acquisition in bioinformatics projects. As stated before, the individual researchers involved in the project often do not develop these rules because of the way that the majority of bioinformatics software is developed to solve a particular research problem as a scientifically oriented solution that the developer is working on rather than for the purpose of setting standards for other researchers as an engineering oriented solution. However, as Contreras notes, the lack of rules may create a situation where bioinformatics researchers are ill prepared for the emergence of standards setting organisations (SSOs). In particular, SSOs may use standards to create a lock out effect for bioinformatics research.[132] This lock out effect may be particularly problematic where certain bioinformatics techniques (such as artificial intelligence and data mining algorithms, as well as data structures) evolve into general purpose technologies. These general purpose technologies will then play an important role across molecular biology research.[133]

To this end, the *sui generis* regime proposed in this thesis might overcome the transaction

---

[130] Dov Greenbaum, 'Patent Sharing in Biotechnology' in Jorge L. Contreras and Meredith Jacob (eds.), *Patent Pledges: Global Perspectives on Patent Law's Private Ordering Frontier* (Edward Elgar Publishing 2017)56–81 59.

[131] Keith E. Maskus and Stephen A. Merrill, *Patent Challenges for Standard-Setting in the Global Economy: Lessons from Information and Communications Technology* (National Academies Press, 2013) 18-9.

[132] Jorge L. Contreras, 'Implementing Procedural Safeguards for the Development of Bioinformatics Interoperability Standards' (2012) 39(2) *Northern Kentucky Law Review* 87–118 113-4.

[133] Iain M. Cockburn, Rebecca Henderson and Scott Stern, above n20, 6 <http://www.nber.org/papers/w24449>.

costs associated with acquiring patents for open innovation. In addition, an innovation patent model focussed on open disclosure and reproducibility of patented technology may bypass the competition law issues that flow from third parties suffering from a lock out when using a standard. Firstly, this partial patent would share the same lowered examination requirements and lifespan of the current innovation patent model, but would require the inventor to completely publish the relevant research results and tests flowing from the application. Secondly, the inventor would be able to submit updated information about the inventors listed with the project at any point of the patent's lifespan, which would encourage the cumulative innovation model observed with software development and other emerging technological disciplines. Finally, this 'partial patent' would place limits on enforcement action that could be taken using the patent.

In addition to the informational benefits of a partial patent model, this model has the benefit of promoting standards setting that de-emphasises patent monetisation arising by only allowing the patents to be used defensively. As Andrew Russell notes, the W3C Foundation's decision to reverse its initial policy on permitting patents on standard technology and encouraging royalty free standards demonstrates that standards setting and royalty free licensing can complement one another.[134] The structuring of these Internet standards SSOs to avoid patent monetisation may therefore provide an example of how future bioinformatics SSOs could be structured to support academic and public-private research.[135] Nevertheless, standards organisation participants remain vulnerable to outside entities that can argue that the standards maintained by the standard organisation that infringe upon their patents.[136] A partial patent model could provide an independent defence to standards litigation by third parties, whilst lowering the transaction costs that a SSO would rely on to obtain such patents.

Recommendation Eight: The feasibility of a *sui generis* form of protection for software should be explored. This *sui generis* right would provide protection for software functionality (in a similar fashion to a patent) and would have lowered examination requirements. However, this *sui generis* regime would also have a shorter period of protection (up to a maximum of ten years). In addition, this *sui generis* regime would permit reverse engineering of software.

---

[134] Andrew L. Russell, 'Constructing Legitimacy: The W3C's Patent Policy' in Laura DeNardis and Michael Zimmer (eds.), *Opening Standards: The Global Politics of Interoperability* (MIT Press 2011) 174.

[135] Jorge L. Contreras, 'Patents and Internet Standards' (Global Commission on Internet Governance Paper No 29, Centre for International Governance Innovation, 15th April 2018) 6-7 <https://www.cigionline.org/publications/patents-and-internet-standards>.

[136] *Rembrandt Wireless Technologies LP v Samsung Electronics Co.* (E.D. Tex., No 2:13-CV-213-JRG, (Feb. 16, 2015)); Jorge L. Contreras, 'When a Stranger Calls: Standards Outsiders and Unencumbered Patents' (2016) 12(3) *Journal of Competition Law and Economics* 507–540 4.

7.5 CONCLUSION

7.5.1 *The Nature of the Bioinformatics Research Environment*

The historical account in Chapter One, as well as the interview results in Chapter Six, have established a number of distinguishing features about bioinformatics research. *First*, bioinformatics research is dominated by academic researchers. A significant amount of bioinformatics software is developed for solving a particular research problem as opposed to developing more generic algorithms. *Secondly*, the vast majority of this bioinformatics research is designed to support upstream (or fundamental) research as opposed to downstream academic research. *Finally*, for these reasons, bioinformatics development evolves very rapidly. These features will influence the intellectual property strategies adopted by bioinformatics research institutes.

In this regard, this thesis has attempted to provide some guidance with respect to these intellectual property strategies by reference to legal precedent and commentary. A comparative approach was chosen because, in addition to being characterised by rapid development, a significant amount of bioinformatics research is conducted on a global scale. Therefore, given the global aspirations of open source development, consistent enforceability would promote the spread of open source development practices. Nevertheless, there is still the potential for the bioinformatics industry to evolve. An increasing number of bioinformatics patents are now held by both overseas and Australian research institutes. Further, as discussed in the Introduction to this thesis, there is continual interest in the downstream research potential of bioinformatics. These competing forces will continue to influence the development of bioinformatics research, both in Australia and overseas.

7.5.2 *The Role of Copyright and Patent Protection in Bioinformatics Development*

Are copyright and patent protection available for bioinformatics software? Chapters Two, Three and Four examined the scope for the protection of bioinformatics algorithms under both copyright law and patent law. Chapter Two established that bioinformatics source code could be copyrightable subject matter in the US, the EU, Australia and New Zealand. Chapter Two also confirmed that open source licences are capable of being enforced as contractual copyright licences in the US and the EU. At present, there is limited case law on the enforcement of open source licences in Australia and New Zealand. However, US and EU precedent might provide a useful guide to answer how open source licences can be effectively enforced. These questions include the extent of protection under open source licence as well as whether licensors can seek non-monetary remedies (namely injunctions) for breaches of open source licences. In particular, whilst Australian courts have not recognised the economic benefits of open source licensing, the operation of cases such as *Jacobsen v Katzer* and the

Welte series might support the economic argument in favour of open source licensing.

In a similar fashion, Chapter Three examined software patent case law on a comparative jurisdictional basis. This doctrinal analysis was necessary due to the longstanding concerns of the impact of software patents on open source development. In recognition of these economic concerns, courts in the US (as well as legislatures in the EU and New Zealand) have attempted to restrict the scope of software patent claiming. Nevertheless, this doctrinal analysis established that there is still a pathway to patent software and algorithms across each of the four jurisdictions considered. Although there are differences in the relevant test, the recurring focus of courts pertains to whether the algorithm in question has some technical effect that elevates it beyond an abstract invention. Accordingly, these tests for patent eligibility would include bioinformatics algorithms that are novel, inventive and have industrial utility. The empirical data presented in this thesis then provides some useful insights about the overlap between open source licensing and copyright and patent licensing.

7.5.3 *Can Copyright and Patent Licensing Impede Open Source Development?*

Chapter Five provided an overview of the patent landscape for academic bioinformatics patents in the US, the EU, Australia and New Zealand. Although academic bioinformatics patents are relatively uncommon (compared to equivalent studies of broader software patents filed by universities), there is a growing number of these patents being filed. However, a review of patent publication pairs (that is, published algorithms which are later patented) suggested that there is not a significant decrease in citation rates following obtaining a patent. Further, the same analysis indicated that the open source licensing of algorithms leads to higher citation rates for the original academic paper.

These patent statistics were coupled with interview based semi-structured interviews as discussed in Chapter Six. The purpose of grounded theory is to inductively generate new theory about sociological phenomena. As a result, there were a number of observations that had been not anticipated at the start of this study. The first observation was the importance of academic recognition as opposed to technology transfer for almost all of the bioinformaticians interviewed for this study. For these interviewees, acknowledgement was identified consistently by interviewees as the primary motivating factor for open source licensing, as it drove ongoing citation. This ongoing citation was measured either by traditional academic citation or alternative citation measures such as the number of downloads or users for a particular software project.[137] Charles Schweik and Robert English have already discussed

[137] There are examples of this motivating factor elsewhere in information systems literature (Kevin Crowston, 'Open Source Technology Development' in William Sims Bainbridge and Mihail C. Roco (eds.), *Handbook of Science and Technology Convergence* (Springer International Publishing 2016)475–486 481-2).

how reputation is a key motivation for developers to contribute to open source projects.[138]

Overall, the findings in this thesis suggest that there are robust open source software communities emerging in bioinformatics research. These communities are at present unhindered by aggressive patent or copyright enforcement. Indeed, this thesis suggests the vast majority of open source bioinformatics development (at least in academia) occurs without exclusive licensing underpinned by patents. Although there is a slow increase in the number of academic bioinformatics patents that are being filed, particularly in the US, interviewees reported that patents did not constitute a strong incentive to engage in bioinformatics development. Instead, the increased recognition that flowed from academic acknowledgements was a far stronger motivating factor for the vast majority of interviewees. For that reason, the majority of interviewees were willing to release their bioinformatics software under an open source licence. This finding confirmed Chapter One's historical account, where the 'gift economies' that underpinned protein sequencing drove nascent open source development practices.

7.5.4 *Recommendations of the Present Study*

Chapter Seven therefore provided recommendations using a combination of private ordering and legislative strategies to support open source bioinformatics research. Specifically, within the context of the Knowledge Commons framework, Chapter Seven recommends introducing improved information and aggregation rules. Improved information rules would be expressed via project and institution level policies for searching for copyright and patent licences (**Recommendation One**). Improved aggregation rules would be supported through a requirement that developers be given greater control over how research software should be licensed (**Recommendation Two**). Chapter Seven also recommended on adopting alternative technology transfer strategies for bioinformatics software beyond exclusive licensing. These technology transfer strategies include dual licensing, the development of bespoke software and the release of software as a service (**Recommendation Three**).

Chapter Seven also recommended supporting these private ordering strategies through legislative reform. First, Chapter Seven recommended replacing the fair dealing exceptions under Australian copyright law with the introduction of a fair use exception into copyright law to act as a pressure valve for the open licensing of emerging technologies (**Recommendation Four**). Secondly, Chapter Seven recommended converging Australian authority on patentable subject matter with the tests elucidated by the UK Court of Appeal in *AerotelMacrossan* to clearly define the degree of technical effect required for a software patent. Adopting this test would raise the standards of patent eligibility and therefore eliminate the problems posed by

[138] Charles M. Schweik and Robert C. English, above n39, 24-34.

expansive software patents without causing confusion (**Recommendation Five**). Thirdly, in order to further protect against aggressive patent enforcement, Chapter Seven recommended strengthening patent exceptions to protect legitimate interoperability (**Recommendation Six**) and to treat open source as a prior use (**Recommendation Seven**). As a final, speculative, recommendation, Chapter Seven discussed investigating a *sui generis* or partial patent model as a means of protecting software functionality. Chapter Seven concluded by examining whether this partial patent model could be used to replace the innovation patent system in Australia (**Recommendation Eight**).

7.5.5 *Future Research*

As with any doctoral study of law, this thesis provides modest recommendations for reform. Further, there are a number of directions that should be addressed for future academic research. However, because the interviews in this thesis were targeted at academic bioinformatics developers, these results may not be entirely reflective of the reasons for engaging in commercial bioinformatics developers may not have the same priorities. Including private sector bioinformaticians in this sample may have also increased the Australian sample size for interviewees (given there are relatively few academic bioinformaticians who have commercialised software). In the alternative, focusing on individual open source projects could be directed towards conducting diagnostic studies of individual open source projects.

Further, as alluded to in Chapter Six, there are other reasons which may militate against the open release of software beyond the existence of formal intellectual property rights. One proprietary right that may influence the open source licensing of software for biomedical research are data protection regulations. Unlike proprietary rights that vest with the developers of software, these regulations provide rights to research subjects. These regimes include the US *Health Insurance Portability and Accountability Act* (HIPAA) and its amending act, the *Health Information Technology for Economic and Clinical Health Act* (HITECH Act). In addition, the EU General Data Protection Regulation (GDPR) regime in the EU establishes new guidelines for the storage and processing of personal data. The GDPR also establishes a regime to allow data subjects to request access and erase their personal data. However, Recital 63 of the GDPR notes that these requests must be subservient to intellectual property protection. This provision may therefore create a mechanism to resolve the conflict between formal intellectual property rights on one hand and data protection law on the other. All of these regimes increase the obligations owed by data collectors, including researchers, to prevent data breaches. The cost of compliance with these regimes is not trivial, and should be considered in a systematic approach alongside other proprietary regimes. Future studies of open source computational biology should therefore consider the effect of unexpected

overlaps between data protection and intellectual property protection regimes.[139]

[139] Madelyn Sanfilippo, Brett Frischmann and Katherine Standburg, 'Privacy as Commons: Case Evaluation Through the Governing Knowledge Commons Framework' (2018) 8() *Journal of Information Policy* 116–166 117-8.

Treaties

Berne Convention for the Protection of Literary and Artistic Works, opened for signature 9th September 1886, 1161 UNTS 30 (entered into force 4th December 1887)

Convention on the Grant of European Patents, opened for signature 5th October 1973, 1065 UNTS 254 (entered into force 7th October 1977)

Marrakesh Agreement Establishing the World Trade Organization, annex IC, The Agreement on Trade Related Aspects of Intellectual Property Rights ('TRIPS Agreement'), opened for signature 15th April 1994, 1867 UNTS 3 (entered into force 1st January 1995)

Paris Convention for the Protection of Industrial Property, opened for signature 14th July 1967, 828 UNTS 305 (entered into force 26th April 1970)

Strasbourg Agreement Concerning the International Patent Classification, opened for signature 21st March 1971, 1160 UNTS 483 (entered into force 7th October 1975)

Treaty on the Functioning of the European Union, opened for signature 7th February 1992, [2009] OJC 115/199 (entered into force 1st November 1993)

WIPO Performances and Phonograms Treaty, opened for signature 20th December 1996, 36 ILM 76 (entered into force 20th May 2002)

World Intellectual Property Organisation Copyright Treaty, opened for signature 20th December 1996, 36 ILM 65 (entered into force 6th March 2002)

European Union Directives

Agreement on a Unified Patent Court [2013] *OJ* C No 175 /56

Council Directive 2009/24/EC of 23 April 2009 on the Legal Protection of Computer Programs [2009] *OJ L* 111/16

Council Directive 91/250 of 14 May 1991 on the Legal Protection of Computer Programs [1991] *OJ* (1 122) No 42

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights of the Information Society [2001] *OJ L* 167/10

Directive 91/250/EEC of 14 May 1991 on the Legal Protection of Computer Programs [1991] *OJ L* 122/42

Directive 96/9/EC on the Legal Protection of Databases [1996] *OJ L* 77/20 No (028/11/EC)

Directive 98/44/EC of the European Parliament and of the Council of 6 July 1998 on the Legal Protection of Biotechnological Inventions [1998] *OJ L* 213/13

Directive of the European Parliament and of the Council on Copyright in the Digital Single Market [2016] *OJ C* 125

Statutes

17 U.S. Code § 101 - Definitions

17 U.S. Code § 106 - Exclusive rights in copyrighted works

35 U.S. Code § 101 - Inventions patentable

35 U.S. Code §273 - Defense to infringement based on prior commercial use

35 U.S. Code § 282 - Presumption of validity; defenses

America Invents Act 2011 (Pub L No 112-129)

American Inventors Protection Act 1999 (Public Law 106-113)

Bayh-Dole Act of 1980 (Patent and Trademarks Act Amendments of 1980)

Belgian Patent Act

Code de la propriété intellectuelle 1992 (French Intellectual Property Code)

Computer Software Copyright Act of 1980 Pub. L. No. 96-517, 94 Stat. 3015 1980

Copyright Act 1968 (Cth)

Copyright Act 1976 (US)

Copyright Act 1994 (NZ)

Copyright Amendment (Computer Programs) Act 1984 (Cth)

Copyright Amendment (Digital Agenda) Act 2000 (Cth)

Copyright, Designs and Patents Act 1988 (UK)

Digital Millennium Copyright Act 17 USC § 1201(a) 1998

Gebrauchsmustergesetz [Utility Model Act] 2009

Patentgesetz [German Patent Act] 1980

Patents Act 1949 (UK)

Patents Act 1977 (UK)

Patents Act 1990 (Cth)

Patents Act 2013 (NZ)

Patents Amendment (Innovation Patents) Act 2000 (Cth)

Statute of Monopolies 1623 (UK)

Cases (Australia, New Zealand and United Kingdom)

Acohs Pty Ltd v Ucorpo Pty Ltd [2012] FCAFC 16

Aerotel Ltd v Telco Holdings Ltd Ors [2006] EWCA Civ 1371

Arrowhead Research Corporation [2016] APO 70

Autodesk Inc v Dyason (No 1) (1992) 173 CLR 330

Autodesk Inc v Dyason (No 2) (1993) 176 CLR 300

Bio-Rad Laboratories, Inc. [2017] APO 24 (3 April 2018)

CA Inc v ISI Pty Ltd (2012) 201 FCR 23

CCOM Pty Ltd v Jiejing Pty Ltd (1996) 21 FCR 260

Chappell & Co Ltd v. Nestle Co Ltd [1960] AC 87

Coastle Recycled Cooking Oils Pty Ltd v Innovative Business Action and Strategies Pty Ltd [2007] NSWSC 831

Commissioner of Patents v RPL Central Pty Ltd [2015] FCAFC 177

Computer Edge Pty Ltd v Apple Computer Inc (1986) 161 CLR 171

Computermate Products (Aust) Pty Ltd v Ozi-Soft Pty Ltd (1988) 12 IPR 487

Coogi Australia Pty Ltd v Hysport International Pty Ltd (1998) 41 IPR 593

D'Arcy v Myriad Genetics Inc (2015) 258 CLR 334

Data Access Corp v Powerflex Services Pty Ltd (1999) 202 CLR 1

Discovery Life Limited [2017] APO 36 (18 July 2017)

Encompass Corporation Pty Ltd v InfoTrack Pty Ltd [2018] FCA 421

Forensic Telecommunications Services Ltd v West Yorkshire Police Anor [2011] EWHC 2892 (Ch)

Fujistu Limited's Application [1997] RPC 608

Gale's Application [1991] RPC 305

Grain Pool of Western Australia v Commonwealth of Australia (2000) 202 CLR 479

Grant v Commissioner of Patents (2006) 154 FCR 62

HTC Europe Co Ltd v Apple Inc [2013] EWCA Civ 451

Hubbard v Vosper [1972] 2 QB 84

Human Genome Sciences Inc v Eli Lilly & Co [2012] RPC 6

IBM v Commissioner of Patents [1991] FCA 625

IceTV Pty Ltd v Nine Network Australia Pty Ltd (2009) 239 CLR 459

Illumina, Inc v Premaitha Health Plc [2017] EWHC 2930 (Pat)

International Business Machines Corp v Computer Imports Ltd [1989] 2 NZLR 395

Jupiters Ltd v Neurizon Pty Ltd [2005] FCAFC 90

Lantana v. Comptroller General of Patents, Design and Trade Marks [2014] EWCA Civ 1463

Lubrizol Corporation & Anor v Esso Petroleum Co Ltd [1998] EWCA Civ 744

Macrossan v Comptroller General of Patents, Designs and Trade Marks 2005 [2006] EWHC 705 (Ch)

Mars, Incorporated v Hill's Pet Nutrition, Inc. [2014] APO 67

Merrill Lynch Inc.'s Application [1988] RPC 1

Microsoft Corp v Commission of the European Communities (T-201/04) [2007] ECR II–3601

Monsanto Co v Stauffer Chemical Co [1985] RPC 515

Monsanto Co v Stauffer Chemical Co (No 1) (1984) 1 NZIPR 518

National Research and Development Corporation v Commissioner of Patents (1959) 102 CLR 252

Newspaper Licensing Agency v Meltwater BV & Ors [2010] EWHC 3099

Research Affiliates LLC v Commissioner of Patents [2014] FCAFC 150

SAS Institute Inc v World Programming Limited [2011] RPC 1

Smith Kline French Laboratories Ltd v Attorney-General (NZ) (1991) 2 NZLR 560

Stephen John Grant [2004] APO 11

Symbian Ltd v Comptroller General of Patents [2008] EWHC 518 (Pat)

Symbian Ltd v Comptroller General of Patents [2009] RPC 1

Telstra Corporation Ltd v Phone Directories Company Pty Ltd [2010] FCA 44

The Court of Edinburgh Napier University (Patent) [2011] UKIntelIP o16411

The Movie Network Channels Pty Ltd v Optus Vision Pty Ltd [2010] NSWCA 111

Todd Martin [2017] APO 33 (7 July 2017)

Trumpet Software Pty Ltd v OzEmail (1996) 34 IPR 481

University of Western Australia v Gray [2009] FCAFC 116

University of Western Australia v Gray (No 20) [2008] FCA 498

Cases (United States)

Alice Corporation v CLS Bank 573 U.S. ¸134 2347 (2014)

Apple Computer, Inc. v Franklin Computer Corp. 714 F. 2d 1240 (3d Cir. 1983)

Apple Computer, Inc. v Microsoft Corp. 799 F.Supp. 1006 (N.D. Cal. 1992)

Apple Computer, Inc. v Microsoft Corp. 35 F.3d 1435 (9th Cir. 1994)

Ariosa Diagnostics, Inc. v Sequenom, Inc. 788 F.3d 1371 (Fed. Cir. 2014)

Association for Molecular Pathology v Myriad Genetics, Inc. 133 S. Ct. 2107 (2013)

Atari Games Corp v Nintendo of America Inc 975 F.2d 832 (Fed. Cir. 1992)

Baker v. Selden 101 U.S. 99 (1880)

Bilski v Kappos 130 S. Ct. 3218 (2010)

Bragg v. Linden Research, Inc. 487 F.Supp. 593 (E.D. Pa. 2007)

Cambridge University Press v Patton 769 F.3d 1232 (11th Cir. 2014)

Campbell v Acuff-Rose Music 510 U.S. 569 (1994)

Carnegie Mellon University v Marvell Technology Group Ltd and Marvell Semiconductor Inc 807 F.3d 1283 (2015)

Computer Associates International Inc v Altai, Inc. 982 F.2d 693 (2d Cir. 1992)

Data Cash Systems, Inc. v JS &a Group, Inc. 480 480 F. Supp. 1063 (1979)

DDR Holdings LLC v Hotels.com L.P. 773 F.3d 1245 (Fed. Cir. 2014)

Diamond v Chakrabarty 447 U.S. 303 (1980)

Diamond v Diehr 450 U.S. 175 (1981)

DNA Twopointo, Inc. v Genome Compiler Corporation (ND. Cal., No No. 3:12-cv-06428-SI, 2013)

Embrex, Inc. v Service Engineering Corp. 216 F.3d 1343 (2000)

Enfish, LLC v Microsoft Corp 822 F.3d. 1327 (Fed. Cir. 2016)

Ex Parte Kelker (Tech Center 1600, No App. No. 2009-004635, (B.P.A.I., September 24 2010))

Feist Publications, Inc. v Rural Telephone Service Co. 499 U.S 340 (1991)

Google, Inc. v Oracle America, Inc.: Supreme Court 135 S. Ct. 2887 (2015)

Gottschalk v Benson 409 U.S. 63 (1972)

Graham v James 144 F.3d 229 (2d Cir. 1998)

IBM v Fujitsu No. 13T-117-0636-85 American Arbitration Ass'n Commercial Arbitration Tribunal 4 (1987)

Illumina, Inc. v Affymetrix, Inc. (Dist. Court. WD Wisconsin, No Nos. 09-cv-277-bbc, 09-cv-665-bbc, 2010)

In re Abele 684 F.2d 902 (C.C.P.A. 1982)

In re Alappat 33 F.3d 1526 (Fed. Cir. 1994)

In re Bilski 545 F.3d 943 (2009)

In re Freeman 573 F.2d 1237 (C.C.P.A. 1978)

In re Walter 618 F.2d 758 (C.C.P.A. 1980)

Jacobsen v Katzer (N.D. Cal., No No. 06-CV-01905 JSW, 2007 WL 2358628, December 2007)

Jacobsen v Katzer 535 F.3d 1373 (Fed. Cir. 2008)

Laboratory Corporation of America Holdings v Metabolite Labs Incorporated 548 U.S. 124 (2006)

Le Roy v Tatham 55 US (14 How.) 156 (1852)

Madey v Duke University 307 F3.d 1351 (2002)

Mayo Collective Services v Prometheus Laboratories 132 S. Ct. 1289 (2012)

MDY Industries, LLC v Blizzard Entertainment 629 F.3d 928 (9th. Cir. 2010)

Oracle America, Inc. v Google Inc. 872 F.Supp.2d 974 (Dist. Court, ND California 2012)

Oracle America, Inc. v Google Inc. 750 F.3d 1339 (Fed. Cir. 2014)

Oracle America, Inc. v Google Inc. (Dist. Court. ND California, No No. C 10-03561 WHA, May 2016)

Oracle America, Inc. v Google Inc. (Fed. Cir., No No. 17-1118, 2018)

O'Reilly v Morse US (15 How.) 62 (1854)

Parker v Flook 437 U.S. 584 (1978)

ProCD, Inc. v Zeidenberg 86 F.3d 1447 (7th Cir. 1996)

Prometheus Laboratories v Mayo Collaborative Ser. 581 F.3d 1336 (2009)

Prometheus Laboratories v Mayo Collaborative Ser. 628 F.3d 1347 (2010)

Regents of the University of California v Broad Institute (Fed. Cir., No No. 2017-1907, 2018)

Rembrandt Wireless Technologies LP v Samsung Electronics Co. (E.D. Tex., No 2:13-CV-213-JRG, (Feb. 16, 2015))

Sega Enterprises Ltd. v Accolade, Inc. 977 F.2d 1510 (9th Cir. 1992)

Sequenom, Inc. v Ariosa Diagnostics, Inc. 136 S. Ct. 2511 (2016)

Specht v Netscape Communications Corp. 306 F.3d 17 (2nd Cir. 2002)

Stanford University v Roche Molecular Systems, Inc 131 S. Ct. 2188 (2011)

State Street Bank Trust Co. v Signature Financial Group Inc. 149 F. 3d 1368 (Fed. Cir. 1998)

Sun Microsystems, Inc. v Microsoft Corp. 21 F.Supp. 1109 (1998)

Sun Microsystems Inc. v Microsoft Corporation (ND. Cal., No No. C 97-20884RMWPVT, 2000 WL 33223397, May 2000)

University City Studios, Inc. v Sony Corp. of America (Betamax) 659 F.2d 963 (9th Cir. 1981)

Whelan Associates, Inc. v Jaslow Dental Laboratory, Inc. 797 F.2d 1222 (3d Cir. 1986)

Wisconsin Alumni Research Foundation v Apple 135 F. Supp. 3d. 865 (2015)

Cases (European Union)

Case C-393/09 Bezpečnostnı softwarová asociace - Svaz softwarové ochrany (BSA) v Ministry of Culture of the Czech Republic [2010] ECR I–13971

Football Dataco Ltd and Others v Yahoo! UK Limited and Others [2010] EUECJ C–604/10

SAS Institute Inc v World Programming Ltd [2013] EUECJ C–406/10

T 0784/06 (23rd June 2010) [2010] 'Genotype determination/BECKMAN')

T 2050/07 (19th February 2013) [2013] 'DNA Mixture Analysis/Perlin')

T0272/95 (23rd October 2002) [2002] 'Relaxin/Howard Florin Institute')

T-1173/97 (1st July 1998) [1997] OJ EPO ('Computer program product/IBM')

T-154/04 (15th November 2006) [2008] O.J. EPO ('DUNS LICENSING ASSOCIATES/Estimating sales activity')

T-172/03 (27th November 2003) [2003] 'RICOH/Order Management')

T-208/84 (15th August 1986) [1987] O.J. EPO ('Vicom/Computer-related invention')

T-38/86 (14th February 1989) [1990] OJ EPO ('IBM/Text Processing')

T-935/97 (4th February 1999) [1997] EPOR ('Computer program product II/IBM')

Cases (German)

adhoc data service GmbH v Buhl Data Service GmbH (WISO Mein Büro), Landgericht (District Court of Bochum) 20th January 2011 reported in 2011 CR 29

Welte v D-Link, Landgericht Frankfurt (District Court of Frankfurt) 22nd September 2006 reported in 2006 CR 729

Welte v Sitecom (Open Source - effectiveness of GPL), Landgericht (District Court München I) 19th May 2004 reported in 2004 MMR 693

Welte v Skype, Landgericht (District Court München I) 12th July 2007 reported in 2007 CR 57

Cases (French)

Tribunal de grande instance de Paris (CA Paris),22nd January 2008 Gaz. Pal. no 22

Cour d'appel de Paris (CA Paris),16th September 2009 RG no 01/24298

Law Reform Commission Reports

Australian Law Reform Commission, *Copyright and the Digital Economy*, Final Report No 122 (November 2013)

Australian Law Reform Commission, *Genes and ingenuity: gene patenting and human health*, No 99 (2004)

Productivity Commission, *Intellectual Property Arrangements*, InquiryReport No 78 (2016)

Productivity Commission, *Intellectual Property Arrangements (Draft)*, Draft Report No 78 (2016)

Ministry of International Trade and Industry, Industrial Structure Council Information Industry Committee, *On the Appropriate Basic Consolidation for Software - Toward the Establishment of Legal Protection of Software*, Interim Reporrt (1983)

Books

John Anderies and Marco Janssen, *Sustaining the commons* (Arizona State University, 2016)

Jonathan Band and Masanobu Katoh, *Interfaces on Trial 2.0* (MIT Press, 2011)

James Bessen and Michael J. Meurer, *Patent Failure: How Judges, Bureaucrats, and Lawyers Put Innovators at Risk* (Princeton University Press, 2009)

Knut Blind, Jakob Edler and Michael Friedewald, *Software Patents: Economic Impacts and Policy Implications* (Edward Elgar Publishing, 2005)

James Boyle, *The Public Domain: Enclosing the Commons of the Mind* (Yale University Press, 2008)

James Boyle and Jennifer Jenkins, *Intellectual Property: Law & the Information Society - Cases & Materials: An Open Casebook: 3rd Edition 2016* (CreateSpace Independent Publishing Platform, 3 editioned, 2016)

J Glenn Brookshear, *Computer Science: An Overview* (Pearson/Addison Wesley, 2009)

David Castle, *The Role of Intellectual Property Rights in Biotechnology Innovation* (Edward Elgar Publishing, 2009)

Nefissa Chakroun, *Patents for development: improved patent information disclosure and access for incremental innovation* (Edward Elgar Publishing, 2016)

Kathy Charmaz, *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis* (SAGE, 2006)

Jorge L. Contreras and A. Jamie Cuticchia, *Bioinformatics Law: Legal Issues for Computational Biology in the Post-genome Era* (American Bar Association, 2013)

Robert M. Cook-Deegan, *The Gene Wars: Science, Politics, and the Human Genome* (W.W. Norton & Company, 1996)

National Research Council, *A New Biology for the 21st Century* (National Academies Press, 2009)

John W. Creswell, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (SAGE Publications, 2012)

Margaret Oakley Dayhoff and National Biomedical Research Foundation, *Atlas of Protein Sequence and Structure. [Vol. 1], [Vol. 1],* (National Biomedical Research Foundation, 1969)

Tom Dedeurwaerdere, *Sustainability Science for Strong Sustainability* (Edward Elgar Publishing, 2014)

Susy Frankel, *Test Tubes for Global Intellectual Property Issues* (Cambridge University Press, 2015)

Susy Frankel and Jessica C. Lai, *Patent Law and Policy* (LexisNexis NZ Limited, 2016)

Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press, 2014)

Miguel Garcıa-Sancho, *Biology, Computing, and the History of Molecular Sequencing: From Proteins to DNA, 1945-2000* (Springer, 2012)

Barney G. Glaser and Anselm L. Strauss, *Discovery of Grounded Theory: Strategies for Qualitative Research* (Routledge, 2017)

William H. Greene, *Econometric Analysis* (Pearson/Prentice Hall, 8th editioned, 2012)

Mark Harvey and Andrew McMeekin, *Public Or Private Economies of Knowledge?: Turbulence in the Biological Sciences* (Edward Elgar, 2007)

Sebastian Haunss, *Conflicts in the Knowledge Society: The Contentious Politics of Intellectual Property* (Cambridge University Press, 2013)

John L. Hennessy and David A. Patterson, *Computer Architecture: A Quantitative Approach* (Elsevier, 2012)

Joseph M. Hilbe, *Negative Binomial Regression* (Cambridge University Press, 2011)

Reto M. Hilty and Sylvie Nérisson, *Balancing Copyright - A Survey of National Approaches* (Springer Science & Business Media, 2012)

Institute of Medicine (US) Forum on Microbial Threats, *The Science and Applications of Synthetic and Systems Biology: Workshop Summary* (National Academies Press (US), 2011)

Thorsten Käseberg, *Intellectual Property, Antitrust and Cumulative Innovation in the EU and the US* (Bloomsbury Publishing, 2012)

Christopher M. Kelty, *Two Bits: The Cultural Significance of Free Software* (Duke University Press, 2008)

Thomas S. Kuhn, *The Structure of Scientific Revolutions: 50th Anniversary Edition* (University of Chicago Press, 1972)

Philip Leith, *Software and Patents in Europe* (Cambridge University Press, 2011)

Tim Lindholm and Frank Yellin, *The Java Virtual Machine Specification* (Addison-Wesley Longman Publishing Co., Inc., 2nded, 1999)

Keith E. Maskus and Stephen A. Merrill, *Patent Challenges for Standard-Setting in the Global Economy: Lessons from Information and Communications Technology* (National Academies Press, 2013)

Henk F. Moed, Wolfgang Glänzel and Ulrich Schmoch, *Handbook of Quantitative Science and Technology Research: The Use of Publication and Patent Statistics in Studies of S&T Systems* (Springer Science & Business Media, 2006)

James E. Monogan III, *Political Analysis Using R* (Springer International Publishing, 2015)

Steven A. Moore, *Pragmatic Sustainability: Theoretical and Practical Tools* (Routledge, 2010)

Dianne Nicol et al., *The Innovation Pool in Biotechnology: The Role of Patents in Facilitating Innovation* (University of Tasmania, 2014)

Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, 1990)

Elinor Ostrom, *Understanding Institutional Diversity* (Princeton University Press, 2005)

Elinor Ostrom, Roy Gardner and James Walker, *Rules, Games, and Common-pool Resources* (University of Michigan Press, 1994)

Amy R. Poteete, Marco Janssen and Elinor Ostrom, *Working Together: Collective Action, the Commons, and Multiple Methods in Practice* (Princeton University Press, 2010)

Viola Prifti, *The Breeder's Exception to Patent Rights: Analysis of Compliance with Article 30 of the TRIPS Agreement* (Springer, 2015)

Alfred Radauer et al., *Study on the Economic Impact of the Utility Model Legislation in Selected Member States: Final Report* (Publications Office, 2015)

Jerome H. Reichman, Tom Dedeurwaerdere and Paul F. Uhlir, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Global Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge University Press, 2016)

Ashwin van Rooijen, *The Software Interface Between Copyright and Competition Law: A Legal Analysis of Interoperability in Computer Programs* (Kluwer Law International, 2010)

Peter H. Salus, *A Quarter Century of UNIX* (Addison-Wesley Publishing Company, 1994)

Jean E. Sammet, *Programming Languages: History and Fundamentals* (Prentice-Hall, Inc., 1969)

Stephen R. Schach, *Classical and object-oriented software engineering with UML and C++* (WCB/McGraw-Hill, 1999)

Charles M. Schweik and Robert C. English, *Internet Success: A Study of Open-Source Software Commons* (MIT Press, 2012)

Suzanne Scotchmer, *Innovation and Incentives* (MIT Press, 2004)

Noam Shemtov, *Beyond the Code: Protection of Non-Textual Features of Software* (Oxford University Press, 2017)

Craig Shimasaki, *Biotechnology Entrepreneurship: Starting, Managing, and Leading Biotech Companies* (Academic Press, 2014)

Johan Söderberg, *Hacking Capitalism: The Free and Open Source Software Movement* (Routledge, 2015)

Hallam Stevens, *Life Out of Sequence: A Data-Driven History of Bioinformatics* (University of Chicago Press, 2013)

Uma Suthersanen, *Utility Models and Innovation in Developing Countries* (International Centre for Trade and Sustainable Development (ICTSD), 2006)

Andrew S. Tanenbaum, *Modern Operating Systems* (Pearson Prentice Hall, 2008)

Robert Tomkowicz, *Intellectual Property Overlaps: Theory, Strategies, and Solutions* (Routledge, 2013)

Frank Vahid and Tony Givargis, *Embedded System Design: a Unified Hardware/Software Introduction* (Wiley, 2002)

Richard Watt, *Handbook on the Economics of Copyright: A Guide for Students and Teachers* (Edward Elgar Publishing, 2014)

Book Chapters

Ross Anderson et al., 'Measuring the Cost of Cybercrime' in *The Economics of Information Security and Privacy* (Springer, Berlin, Heidelberg 2013)265

Teresa. K. Attwood et al., 'Concepts, historical milestones and the central place of bioinformatics in modern biology: a European perspective' in *Bioinformatics - trends and methodologies* (InTech 2011)

David B. Audretsch and Devrim Göktepe-Hultén, 'University Patenting in Europe: Does Faculty Ownership of Intellectual Property Impede University Technology Transfer?' in Albert N. Link, Donald S. Siegel and Mike Wright (eds.), *The Chicago Handbook of University Technology Transfer and Academic Entrepreneurship* (University of Chicago Press 2015)

Yochai Benkler, 'Between Spanish Huertas and the Open Road: A Tale of Two Commons?' in Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press 2014)69–98

Yochai Benkler, 'Peer Production and Cooperation' in Johannes M. Bauer and Michael Latzer (eds.), *Handbook on the Economics of the Internet* (2016)

Nicolas Binctin, 'The French Copyright Law Opens Its Arms to the FOSS' in Axel Metzger (ed.), *Free and Open Source Software (FOSS) and other Alternative License Models* (Springer, Cham 2016)477–493

John Braithwaite, Dianne Nicol and Janet Hope, 'Regulatory Captialism, Business Models and the Knowledge Economy' in John Braithwaite (ed.), *Regulatory Capitalism: How it Works, Ideas for Making it Work Better* (Edward Elgar 2008)

Tania M. Bubela et al., 'Governance of Biomedical Research Commons to Advance Clinical Translation: Lessons from the Mouse Model Community' in Katherine J. Strandburg, Brett M. Frischmann and Michael J. Madison (eds.), *Governing Medical Knowledge Commons* (Cambridge University Press 2017)222

Vincent J. Carey and Victoria Stodden, 'Reproducible Research Concepts and Tools for Cancer Bioinformatics' in Michael F. Ochs, John T. Casagrande and Ramana V. Davuluri (eds.), *Biomedical Informatics for Cancer Research* (Springer US 2010)149–175

Daniel H. Cole, 'Learning from Lin: Lessons and Cautions from the Natural Commons for the Knowledge Commons' in Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press 2014)45–68

Giorgio Fabio Colombo and Matteo Dragoni, 'The Legal Protection of Software in Japan—An Original Model?' in Giuseppe Bellantuono and Fabiano Teodoro Lara (eds.), *Law, Development and Innovation* (Springer 2016)67–88

Jorge L. Contreras, 'Leviathan in the Commons' in Katherine J. Strandburg, Brett M. Frischmann and Michael J. Madison (eds.), *Governing Medical Knowledge Commons* (Cambridge University Press 2017)

Kevin Crowston, 'Open Source Technology Development' in William Sims Bainbridge and Mihail C. Roco (eds.), *Handbook of Science and Technology Convergence* (Springer International Publishing 2016)475–486

Mihály Ficsor, 'Collective Management of Copyright and Related Rights from the Viewpoint of International Norms and the Acquis Communautaire' in Daniel J. Gervais (ed.), *Collective Management of Copyright and Related Rights* (Kluwer Law International 2010)

Brett Frischmann, Michael Madison and Katherine Strandburg, 'Governing Knowledge Commons – Introduction & Chapter 1' in Brett M. Frischman, Michael Madison and Katherine Strandburg (eds.), *Governing Knowledge Commons* (2014)

Shubha Ghosh, 'How to Build a Commons: Is Intellectual Property Constrictive, Facilitating, or Irrelevant?' in *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press 2007)209–245

Wendy J. Gordon, 'The Fair Use Doctrine: Markets, Market Failure and Rights of Use' in Richard Watt (ed.), *Handbook on the Economics of Copyright: A Guide for Students and Teachers* (Edward Elgar Publishing 2014)

Guido Governatori et al., 'One License to Compose Them All' in Harith Alani et al. (eds.), *The Semantic Web – ISWC 2013* (Springer Berlin Heidelberg 2013)151–166

Stuart JH Graham and David C. Mowery, 'Intellectual property protection in the US software industry' in Wesley M. Cohen and Stephen A. Merrill (eds.), *Patents in the Knowledge-based Economy* (National Research COuncil 2003)219231

Dov Greenbaum, 'Patent Sharing in Biotechnology' in Jorge L. Contreras and Meredith Jacob (eds.), *Patent Pledges: Global Perspectives on Patent Law's Private Ordering Frontier* (Edward Elgar Publishing 2017)56–81

Christian Handke, Lucie Guibault and Joan-Josep Vallbé, 'Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research' in *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science: Scale, Openness and Trust: proceedings of the 19th International Conference on Electronic Publishing* (2015)120

Charlotte Hess and Elinor Ostrom, 'A Framework for Analyzing the Knowledge Commons' in Charlotte Hess and Elinor Ostrom (eds.), *Understanding Knowledge as a Commons: from Theory to Practice* (MIT Press 2007)

A. J. G. Hey and A. E. Trefethen, 'The Data Deluge: An e-Science Perspective' in F. Berman, G. C. Fox and A. J. G. Hey (eds.), *Grid Computing - Making the Global Infrastructure a Reality* (Wiley and Sons 2003)809–824

Janet Hope, 'An Introduction to Open Source Biotechnology' in Christopher Arup and William van Caenegem (eds.), *Intellectual Property Policy Reform: Fostering Innovation and Development* (Edward Elgar 2009)

Helena Howe, Jonathan Griffiths and S?verine Dusollier, 'The Commons as a Reverse Intellectual Property – From Exclusivity to Inclusivity' in *Concepts of Property in Intellectual Property Law* (Cambridge University Press 2013)258

Gregory J. Kirsch and Charlie F. Brown, ‘Software patents’ in Jorge L. Contreras and A. Jamie Cuticchia (eds.), *Bioinformatics Law: Legal Issues for Computational Biology in the Post-genome Era* (American Bar Association 2013)

Larry L Kiser and Elinor Ostrom, ‘The Three Worlds of Action: A Metatheoretical Synthesis of Institutional Approaches’ in Michael Dean McGinnis (ed.), *Polycentric Games and Institutions: Readings from the Workshop in Political Theory and Policy Analysis* (University of Michigan Press 2000)

Jessica C. Lai, ‘The Nebulous "Invention": From "Idea and Embodiment" to "Idea/Embodiment and Observable Physical Effects?"’ in Antoinette Maget Dominicé and Jessica C. Lai (eds.), *Intellectual Property and Access to Im/material Goods* (Edward Elgar Publishing 2016)121

Peter Lee, ‘Centralization, Fragmentation, and Replication in the Genomic Data Commons’ in Katherine J. Strandburg, Brett M. Frischmann and Michael J. Madison (eds.), *Governing Medical Knowledge Commons* (Cambridge University Press 2017)

Sabina Leonelli, ‘Valuing data in postgenomic biology : how data donation and curation practices challenge the scientific publication system’ in Sarah S. Richardson and Hallam Stevens (eds.), *Postgenomics : perspectives on biology after the genome* (Duke University Press 2015)126–149

‘Empirical Research in International Management: A Critique and Future Agenda’ in Rebecca Marschan-Piekkari and Catherine Welch (eds.), *Handbook of Qualitative Research Methods for International Business* (Edward Elgar 2004)25

Luke McDonagh, ‘Copyright, Contract and FOSS’ in Noam Shemtov and Ian Walden (eds.), *Free and Open Source Software: Policy, Law and Practice* (Oxford University Press 2013)71–108

Axel Metzger and Stefan Hennigs, ‘General Report’ in Axel Metzger (ed.), *Free and Open Source Software (FOSS) and other Alternative License Models: A Comparative Analysis* (Springer 2015)

Timo Minssen, 'Patenting Human Genes in Europe- And how it compares to the US and Australia' in Duncan Matthews and Herbert Zech (eds.), *Research Handbook on Intellectual Property and the Life Sciences* (Edward Elgar Publishing Limited 2017)26

Elinor Ostrom, 'Policies that Crowd out Reciprocity and Collective Action' in Herbert Gintis (ed.), *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life* (MIT Press 1st Editioned, 2005)253

Elinor Ostrom and Sue E. S. Crawford, 'A Grammar of Institutions' in Elinor Ostrom (ed.), *Understanding Institutional Diversity* (Princeton University Press 2009)137–174

Elinor Ostrom and Sue E. S. Crawford, 'Classifying Rules' in Elinor Ostrom (ed.), *Understanding Institutional Diversity* (Princeton University Press 2009)168

Kenneth A. Oye and Rachel Wellhausen, 'The Intellectual Commons and Property in Synthetic Biology' in *Synthetic Biology* (Springer, Dordrecht 2009)121–140

Kylie Pappalardo and Brian Fitzgerald, 'Copyright, Fair Use and the Australian Constitution' in Brian Fitzgerald and John Gilchrist (eds.), *Copyright Perspectives: Past, Present and Prospect* (Springer 2015)125

Andrew L. Russell, 'Constructing Legitimacy: The W3C's Patent Policy' in Laura DeNardis and Michael Zimmer (eds.), *Opening Standards: The Global Politics of Interoperability* (MIT Press 2011)

Pamela Samuelson, 'A Square Peg in a Round Hole? Copyright Protection for Computer Programs' in Brad Sherman and Wiseman (eds.), *Copyright and the Challenge of the New* (Kluwer Law International 2012)

Noam Shemtov, 'Circumventing the Idea/Expression Dichotomy: The Use of Copyright, Technology and Contract to Deny Access to Ideas' in Guido Westkamp (ed.), *Emerging Issues in Intellectual Property: Trade, Technology and Market Freedom : Essays in Honour of Herchel Smith* (Edward Elgar Publishing 2007)

Victoria Stodden, 'Enabling Reproducibility in Big Data Research: Balancing Confidentiality and Scientific Transparency' in *Privacy, Big Data and the Public Good* (Cambridge University Press 2014)112

Victoria Stodden, 'What Computational Scientists Need to Know about Intellectual Property Law' in *Implementing Reproducible Research* (Chapman and Hall/CRC 2014)

Geertrui Van Overwalle, 'Governing Genomic Data: Plea for an "Open Commons"' in Brett M. Frischmann, Michael J. Madison and Katherine J. Strandburg (eds.), *Governing Knowledge Commons* (Oxford University Press 2014)137–154

Geertrui Van Overwalle et al., 'Inventing Inclusive Patents. From Old to New Open Innovation' in *Kritika* (2015)1206–277

James G.S. Wilson, 'On the Value of the Intellectual Commons' in Annabelle Lever (ed.), *New Frontiers in the Philosophy of Intellectual Property* (Cambridge University Press 2012)18122–139

Esther. van Zimmeren and Geertrui Van Overwalle, 'False Sense of Security Offered by Zero-Price Liability Rules? Research Exceptions in the US, Europe and Japan in an Open Innovation Context' in Ruth Okediji and Margo Bagly (eds.), *Global Perspectives on Patent Law* (Cambridge University Press 2011)379

Articles

Paul S. Adler and Seok-Woo Kwon, 'Social Capital: Prospects for a New Concept' (2002) 27(1) *The Academy of Management Review* 17–40

Nitin Aggarwal and Eric A. Walden, 'Intellectual Property Bundle (IPB) theory: Managing Transaction Costs in Technology Development Through Network Governance' (2009) 48(1) *Decision Support Systems* 23–32

Isabella Alexander, 'Manacles upon Science: Re-Evaluating Copyright in Informational Works in Light of 18th Century Case Law' (2014) 38(2) *Melbourne University Law Review* 317–361

Sarah E. Ali-Khan et al., 'Defining Success in Open Science' (2018) 2(2) *MNI Open Research* 1

Simone Aliprandi, 'Open licensing and databases' (2012) 4(1) *International Free and Open Source Software Law Review* 5–18

John R. Allison and Mark A. Lemley, 'Who's Patenting What - An Empirical Exploration of Patent Prosecution' (2000) 53(6) *Vanderbilt Law Review* 2099–2174

Stephen F. Altschul et al., 'Basic local alignment search tool' (1990) 215(3) *Journal of Molecular Biology* 403–410

John Anderies, Marco Janssen and Edella Schlager, 'Institutions and the performance of coupled infrastructure systems' (2016) 10(2) *International Journal of the Commons*

Birgitte Andersen and Federica Rossi, 'Inefficiencies in markets for intellectual property rights: experiences of academic and public research institutions' (2012) 30(1) *Prometheus* 5–27

Christopher Anderson, 'US patent application stirs up gene hunters' (1991) 353(6344) *Nature* 485–486

Ross Anderson and Tyler Moore, 'Information Security: Where Computer Science, Economics and Psychology Meet' (2009) 367(1898) *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 2717–2727

J. Andreu-Perez et al., 'Big Data for Health' (2015) 19(4) *IEEE Journal of Biomedical and Health Informatics* 1193–1208

Rachel A. Ankeny, 'The Natural History of Caenorhabditis elegans Research' (2001) 2(6) *Nature Reviews Genetics* 474–479

Rachel A. Ankeny and Sabina Leonelli, 'Repertoires: A post-Kuhnian perspective on scientific change and collaborative research' (2016) 60(Supplement C) *Studies in History and Philosophy of Science Part A* 18–28

Wilhelm J. Ansorge, 'Next-generation DNA sequencing techniques' (2009) 25(4) *New biotechnology* 195–203

Francesco Paolo Appio, Antonella Martini and Gualtiero Fantoni, 'The light and shade of knowledge recombination: Insights from a general-purpose technology' (2017) 125(12) *Technological Forecasting and Social Change* 154–165

Christine Årdal and John-Arne Røttingen, 'Open Source Drug Discovery in Practice: A Case Study' (2012) 6(9) *PLoS Neglected Tropical Diseases* e1827

Jalayne J. Arias, Genevieve Pham-Kanter and Eric G. Campbell, 'The Growth and Gaps of Genetic Data Sharing Policies in the United States' (2015) 2(1) *Journal of Law and the Biosciences* 56–68

Michael Armbrust et al., 'A View of Cloud Computing' (2010) 53(4) *Communications of the ACM* 50–58

Ashish Arora, Lee G. Branstetter and Matej Drev, 'Going Soft: How the Rise of Software-Based Innovation Led to the Decline of Japan's IT Industry and the Resurgence of Silicon Valley' (2012) 95(3) *Review of Economics and Statistics* 757–775

Clark D. Asay, 'Copyright's Technological Interdependencies' (2015) 18(2) *Stanford Technology Law Review* 189–497

Clark D. Asay, 'Enabling Patentless Innovation' (2014) 74(3) *Maryland Law Review* 431

Clark D. Asay, 'Software's Copyright Anticommons' (2016) 66(2) *Emory Law Journal* 265–332

Joaquın M. Azagra-Caro, Luis Plaza-Gómez and Ana Romero-de-Pablos, 'The origin of public research organisation patents: an economic approach' (2007) 16(4) *Research Evaluation* 271–282

Pierre Azoulay, Waverly Ding and Toby Stuart, 'The Impact of Academic Patenting on the Rate, Quality and Direction of (Public) Research Output' (2009) 57(4) *The Journal of Industrial Economics* 637–676

Jacopo Baggio et al., 'Explaining Success and Failure in The Commons: The Configural Nature of Ostrom's Institutional Design Principles' (2016) 10(2) *International Journal of the Commons*

Judith Bannister, 'Open Government: From Crown Copyright to the Creative Commons and Culture Change Forum: Intellectual Property Law in Australia' (2011) 34(3) *University of New South Wales Law Journal* 1080–1103

Andrew Bartlett, Jamie Lewis and Matthew L. Williams, 'Generations of Interdisciplinarity in Bioinformatics' (2016) 35(2) *New Genetics and Society* 1–24

Andrew Bartlett et al., 'The Locus of Legitimate Interpretation in Big Data Sciences: Lessons for Computational Social Science rom -Omic Biology and High-Energy Physics' (2018) 5(1) *Big Data & Society* 2053951718768831

Xavier Basurto et al., 'A Systematic Approach to Institutional Analysis: Applying Crawford and Ostrom's Grammar' (2010) 63(3) *Political Research Quarterly* 523–537

Brendon Beheshti, 'Getting beyond Abstract Confusion: How the United Kingdom's Jurisprudence Can Aid in Developing an Analytic Framework for Patent-Eligibility in Light of Alice v. CLS Bank' (2014) 10(2) *Washington Journal of Law, Technology & Arts* 137–152

Paul Berg and Janet E. Mertz, 'Personal Reflections on the Origins and Emergence of Recombinant DNA Technology' (2010) 184(1) *Genetics* 9–17

Florian Berger, Knut Blind and Nikolaus Thumm, 'Filing Behaviour Regarding Essential Patents in Industry Standards' (2012) 41(1) *Research Policy* 216–225

Gaia Bernstein, 'The Rise of the End User in Patent Litigation' (2014) 55(5) *Boston College Law Review* 1443–1500

James Bessen and Robert M. Hunt, 'An Empirical Look at Software Patents' (2007) 16(1) *Journal of Economics & Management Strategy* 157–189

Krithika Bhuvaneshwar et al., 'A Case Study for Cloud Based High Throughput Analysis of NGS Data Using the Globus Genomics System' (2015) 13(Supplement C) *Computational and Structural Biotechnology Journal* 64–74

Mario Biagioli, 'Patent Republic: Representing Inventions, Constructing Rights and Authors' (2006) 73(4) *Social Research: An International Quarterly* 1129–1172

Curt Blake and Joseph Probst, 'Loaded Question: Examining Loadable Kernel Modules under the General Public License v2' (2011) 7(3) *Washington Journal of Law, Technology & Arts* 265–294

Andrea Bonaccorsi, Jane Calvert and Pierre-Benoit Joly, 'From Protecting Texts to Protecting Objects in Biotechnology and Software: A Tale of Changes of Ontological Assumptions in Intellectual Property Protection' (2011) 40(4) *Economy and Society* 611–639

Christine L. Borgman, 'The Conundrum of Sharing Research Data' (2012) 63(6) *Journal of the American Society for Information Science and Technology* 1059–1078

Lutz Bornmann and Hans-Dieter Daniel, 'Selecting Scientific Excellence through Committee Peer Review - A Citation Analysis of Publications Previously Published to Approval or Rejection of Post-Doctoral Research Fellowship Applicants' (2006) 68(3) *Scientometrics* 427–440

Lutz Bornmann and Werner Marx, 'The h index as a Research Performance Indicator' (2011) 37(3) *European Science Editing* 77–80

Kevin J. Boudreau and Karim R. Lakhani, "Open" Disclosure of Innovations, Incentives and Follow-On Reuse: Theory on Processes of Cumulative Innovation and a Field Experiment in Computational Biology' (2015) 44(1) *Research Policy* 4–19

Maarten Boudry and Massimo Pigliucci, 'The Mismeasure of Machine: Synthetic Biology and the Trouble with Engineering Metaphors' (2013) 44(4) *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 660–668

Glenn A. Bowen, 'Grounded Theory and Sensitizing Concepts' (2006) 5(3) *International Journal of Qualitative Methods* 12–23

Virginia Braun and Victoria Clarke, 'Using thematic analysis in psychology' (2006) 3(2) *Qualitative Research in Psychology* 77–101

Stephen Breyer, 'The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs' (1970) 84(2) *Harvard Law Review* 281–351

C. L. Brown, 'AT&T and the Consent Decree' (1983) 7(2) *Telecommunications Policy* 91–95

Tania M. Bubela and Timothy Caulfield, 'Role and Reality: Technology Transfer at Canadian Universities' (2010) 28(9) *Trends in Biotechnology* 447–451

Tania Bubela, Jenilee Guebert and Amrita Mishra, 'Use and Misuse of Material Transfer Agreements: Lessons in Proportionality from Research, Repositories, and Litigation' (2015) 13(2) *PLoS Biol* e1002060

Tania Bubela et al., 'Managing Intellectual Property to Promote Pre-Competitive Research: The Mouse as a Model for Constructing a Robust Research Commons' (2012) 22(1) *Journal of Law, Information and Science* 98

Tania Bubela et al., 'Patent Landscaping for Life Sciences Innovation: Toward Consistent and Transparent Practices' (2013) 31(3) *Nature Biotechnology* 202–206

Jochen Bühling, 'The Protection of Computer-Implemented Inventions – An Analysis of the Latest Case Law of the German Federal Court of Justice' (2013) 8(12) *Journal of Intellectual Property Law & Practice* 957–960

Michael Callaham, Robert L. Wears and Ellen Weber, 'Journal Prestige, Publication Bias, and Other Characteristics Associated With Citation of Published Studies in Peer-Reviewed Journals' (2002) 287(21) *JAMA* 2847–2850

Jane Calvert, 'The Commodification of Emergence: Systems Biology, Synthetic Biology and Intellectual Property' (2008) 3(04) *BioSocieties* 383–398

Jane Calvert and Joan H. Fujimura, 'Calculating life? Duelling Discourses in Interdisciplinary Systems Biology' (2011) 42(2) *Studies in History and Philosophy of Biological and Biomedical Sciences* 155–163

Mercedes Campi and Alessandro Nuvolari, 'Intellectual Property Protection in Plant Varieties: A Worldwide Index (1961–2011)' (2015) 44(4) *Research Policy* 951–964

Dr. Jose J. Gonzalez de Alaiza Cardona, 'Open Source, Free Software, and Contractual Issues' (2007) 15(2) *Texas Intellectual Property Law Journal* 157

Bryan Carson, 'Legally Speaking – The Legality of Open Source Part II: Jacobsen v. Katzer' (2013) 21(1) *Against the Grain*

David W. Cash et al., 'Scale and Cross-Scale Dynamics: Governance and Information in a Multilevel World' (2006) 11(2) *Ecology and Society*

Henry Chesbrough and Marshall Van Alstyne, 'Permissionless Innovation' (2015) 58(8) *Communications of the ACM* 24–26

Colleen V. Chien, 'Of Trolls, Davids, Goliaths, and Kings: Narratives and Evidence in the Litigation of High-Tech Patents Frontiers in Empirical Patent Law Scholarship' (2008) 87(5) *North Carolina Law Review* 1571–1616

Elliot J. Chikofsky and James H. Cross, 'Reverse engineering and design recovery: a taxonomy' (1990) 7(1) *IEEE Software* 13–17

Peter A. Chow-White and Miguel Garcıa-Sancho, 'Bidirectional Shaping and Spaces of Convergence: Interactions between Biology and Computing from the First DNA Sequencers to Global Genome Databases' (2012) 37(1) *Science, Technology, & Human Values* 124–164

Julie E. Cohen and Mark A. Lemley, 'Patent Scope and Innovation in the Software Industry' (2001) 89(1) *California Law Review* 1–57

Francis S. Collins, Michael Morgan and Aristides Patrinos, 'The Human Genome Project: Lessons from Large-Scale Biology' (2003) 300(5617) *Science* 286–290

Jorge L. Contreras, 'Bermuda's Legacy: Policy, Patents, and the Design of the Genome Commons' (2011) 12(1) *Minnesota Journal of Law, Science & Technology* 61

Jorge L. Contreras, 'Implementing Procedural Safeguards for the Development of Bioinformatics Interoperability Standards' (2012) 39(2) *Northern Kentucky Law Review* 87–118

Jorge L. Contreras, 'Patent Pledges' (2015) 47(3) *Arizona State Law Journal* 543–608

Jorge L. Contreras, 'When a Stranger Calls: Standards Outsiders and Unencumbered Patents' (2016) 12(3) *Journal of Competition Law and Economics* 507–540

Jorge L. Contreras, Aris Floratos and Arthur L. Holden, 'The International Serious Adverse Events Consortium's Data Sharing Model' (2013) 31(1) *Nature Biotechnology* 17–19

Charles E. Cook et al., 'The European Bioinformatics Institute in 2017: Data Coordination and Integration' (2018) 46(D1) *Nucleic Acids Research* D21–D29

Robert Cook-Deegan, Rachel A. Ankeny and Kathryn Maxson Jones, 'Sharing Data to Build a Medical Information Commons: From Bermuda to the Global Alliance' (2017) 18(1) *Annual Review of Genomics and Human Genetics* 389

Philip Cooke, 'Biosciences and the Rise of Regional Science Policy' (2004) 31(3) *Science and Public Policy* 185–197

Philip Cooke, 'European Asymmetries: A Comparative Analysis of German and UK Biotechnology Clusters' (2007) 34(7) *Science and Public Policy* 454–474

Philip Cooke, 'Life Sciences Clusters and Regional Science Policy' (2004) 41(5-6) *Urban Studies* 1113–1131

Susan Corbett, 'Creative Commons Licences, the Copyright Regime and the Online Community: Is there a Fatal Disconnect?' (2011) 74(4) *The Modern Law Review* 503–531

Susan Corbett, ‘What if Object Code Had Been Excluded from Protection as a Literary Work in Copyright Law - A New Zealand Perspective’ (2008) *Michigan State Law Review* 173–198

Juliet M. Corbin and Anselm Strauss, ‘Grounded Theory Research: Procedures, Canons, and Evaluative Criteria’ (1990) 13(1) *Qualitative Sociology* 3–21

Rodrigo Costas, Zohreh Zahedi and Paul Wouters, ‘Do “Altmetrics” Correlate with Citations? Extensive Comparison of Altmetric Indicators with Citations From a Multidisciplinary Perspective’ (2015) 66(10) *Journal of the Association for Information Science and Technology* 2003–2019

Thomas F. Cotter, ‘Fair Use and Copyright Overenforcement’ (2007) 93(4) *Iowa Law Review* 1271–1318

Sue E. S. Crawford and Elinor Ostrom, ‘A Grammar of Institutions’ (1995) 89(3) *American Political Science Review* 582–600

Paola Criscuolo, ‘The ’Home Advantage’ Effect and Patent Families. A Comparison of OECD Triadic Patents, the USPTO and the EPO’ (2006) 66(1) *Scientometrics* 23–41

Dennis Crouch and Robert P. Merges, ‘Operating Efficiently Post-Bilski by Ordering Patent Doctrine Decision-Making’ (2010) 25(4) *Berkeley Technology Law Journal* 1673–1692

Kevin Crowston et al., ‘Free/Libre Open-Source Software Development: What We Know and What We Do Not Know’ (2012) 44(2) *ACM Computing Surveys* 7–7:35

Peter A. Cummings, ‘From Germany to Australia: Opportunity for a Second Tier Patent System in the United States’ (2009) 18(2) *Michigan State University College of Law Journal of International Law* 297–322

James E. Daily, ‘Alice’s Aftermath: Changes in Patentee Behavior since Alice v. CLS Bank Bridging the Gap between the Federal Courts and the United States Patent & Trademark Office: The Journal of Science and Technology Law Symposium’ (2017) 23() *Boston University Journal of Science and Technology Law* 284–303

Hendrik P. van Dalen and Kéne Henkens, ‘Signals in Science - On the Importance of Signaling in Gaining Attention in Science’ (2005) 64(2) *Scientometrics* 209–233

Kenneth W. Dam, 'Some Economic Considerations in the Intellectual Property Protection of Software' (1995) 24(2) *Journal of Legal Studies* 321–378

Antoine Danchin, 'Information of the Chassis and Information of the Program in Synthetic Cells' (2009) 3(1-4) *Systems and Synthetic Biology* 125–134

Partha Dasgupta and Paul A. David, 'Toward a New Economics of Science' (1994) 23(5) *Research Policy* 487–521

Paul A. David, 'Understanding the Emergence of 'Open Science' Institutions: Functionalist Economics in Historical Context' (2004) 13(4) *Industrial and Corporate Change* 571–589

Philip M. Davis et al., 'Open Access Publishing, Article Downloads, and Citations: Randomised Controlled Trial' (2008) [337] *BMJ* a568

Primavera De Filippi and Miguel Said Vieira, 'The Commodification of Information Commons: The Case of Cloud Computing' (2014) 16(1) *Columbia Science and Technology Law Review* 102–143

Hannah Deakin and Kelly Wakefield, 'Skype Interviewing: Reflections of Two PhD Researchers' (2014) 14(5) *Qualitative Research* 603–616

Peter J. Denning, 'The Science of Computing: The ARPANET after Twenty Years' (1989) 77(6) *American Scientist* 530–534

Chris Dent, 'The TRIPS agreement and an experimental use exception for 'research tools'' (2011) 44(1) *Australian Economic Review* 73–78

Jonathan M. J. Derry et al., 'Developing Predictive Molecular Maps of Human Disease Through Community-Based Modeling' (2012) 44(2) *Nature Genetics* 127–130

Lothar Determann, 'Dangerous Liaisons - Software Combinations as Derivative Works - Distribution, Installation, and Execution of Linked Programs under Copyright Law, Commercial Licenses, and the GPL' (2006) 21(4) *Berkeley Technology Law Journal* 1421

Lothar Determann, 'Software Copyright's Oracle from the Cloud' (2015) 30(1) *Berkeley Technology Law Journal* 161

Lothar Determann and David Nimmer, 'Software Copyright's Oracle from the Cloud' (2014) 30(1) *Berkeley Technology Law Journal* 161

Erwin L. van Dijk et al., 'Ten Years of Next-Generation Sequencing Technology' (2014) 30(9) *Trends in Genetics* 418–426

Graeme B. Dinwoodie and Rochelle Cooper Dreyfuss, 'Brexit and IP: The Great Unraveling' (2017) 39() *Cardozo Law Review* 967–996

Laurence Diver, 'Would the Current Ambiguities Within the Legal Protection of Software be Solved by the Creation of a Sui Generis Property Right for Computer Programs?' (2008) 3(2) *Journal of Intellectual Property Law & Practice* 125–138

Gerardo Con Dıaz, 'The Text in the Machine: American Copyright Law and the Many Natures of Software, 1974–1978' (2016) 57(4) *Technology and Culture* 753–779

Antony W. Dnes, 'Should the UK Move to a Fair-Use Copyright Exception?' (2013) 44(4) *IIC - International Review of Intellectual Property and Competition Law* 418–444

Mark Dodgson et al., 'Systems Thinking, Market Failure, and the Development of Innovation Policy: The Case of Australia' (2011) 40(9) *Research Policy* 1145–1156

Peter Van Dongen et al., 'The Relationships Between University IP Regimes, Scientists' Motivations and Their Engagement with Research Commercialisation in Europe' (2017) 8(2) *European Journal of Law and Technology*

Peter van Dongen, Hester Tak and Eric Claassen, 'Policies and patenting to stimulate the biotechnology sector: evidence from The Netherlands' (2018) 46(1) *Science and Public Policy* 136–147

Eli Dourado and Alex Tabarrok, 'Public Choice Perspectives on Intellectual Property' (2015) 163(1) *Public Choice* 129–151

Robert R. Downs et al., 'Community Recommendations for Sustainable Scientific Software' (2015) 3(1) *Journal of Open Research Software*

Rochelle Cooper Dreyfuss, 'Reconsidering Experimental Use' (2016) 50(4) *Akron Law Review* 699–724

Rochelle Cooper Dreyfuss, 'The Federal Circuit: A Case Study in Specialized Courts' (1989) 64(1) *New York University Law Review* 1

Claire T. Driscoll, 'NIH Data and Resource Sharing, Data Release and Intellectual Property Policies for Genomics Community Resource Projects' (2005) 15(1) *Expert Opinion on Therapeutic Patents* 1–8

Geraint Duck et al., 'A Survey of Bioinformatics Database and Software Usage through Mining the Literature' (2016) 11(6) *PLOS ONE* e0157989

Peter Eckersley et al., 'Neuroscience Data and Tool Sharing' (2003) 1(2) *Neuroinformatics* 149–165

Pehr Edman and Gregory Begg, 'A Protein Sequenator' (1967) 1(1) *European Journal of Biochemistry* 80–91

Rebecca S. Eisenberg, 'Diagnostics Need Not Apply' (2015) 21(1) *Boston University Journal of Science and Technology Law* 256–286

Rebecca S. Eisenberg, 'Genes, Patents, and Product Development' (1992) 257(5072) *Science* 903–908

Rebecca S. Eisenberg, 'Noncompliance, Nonenforcement, Nonproblem - Rethinking the Anticommons in Biomedical Research Patent Law in Perspective' (2008) 45(4) *Houston Law Review* 1059–1100

Rebecca S. Eisenberg, 'Patents and the Progress of Science: Exclusive Rights and Experimental Use' (1989) 56(3) *University of Chicago Law Review* 1017

Rebecca S. Eisenberg, 'Prometheus Rebound: Diagnostics, Nature, and Mathematical Algorithms' (2013) 122(1) *Yale Law Journal Online* 341–837

Rebecca S. Eisenberg, 'Proprietary Rights and the Norms of Science in Biotechnology Research' (1987) 97(2) *Yale Law Journal* 177–231

Rebecca S. Eisenberg and Richard R. Nelson, 'Public vs. Proprietary Science: A Fruitful Tension?' (2002) 77(12, Part 2) *Academic Medicine* 1392–1399

Kathleen M. Eisenhardt, 'Building Theories from Case Study Research' (1989) 14(4) *Academy of Management Review* 532–550

Ezieddin Elmahjub and Nicolas P. Suzor, 'Fair Use and Fairness in Copyright: A Distributive Justice Perspective on Users' Rights' (2017) 43(1) *Monash University Law Review* 274–298

Michael B. Elowitz and Stanislas Leibler, 'A Synthetic Oscillatory Network of Transcriptional Regulators' (2000) 403(6767) *Nature* 335–338

Drew Endy, 'Foundations for Engineering Biology' (2005) 438(7067) *Nature* 449–453

Sebastian v. Engelhardt and Andreas Freytag, 'Institutions, culture, and open source' (2013) 95() *Journal of Economic Behavior & Organization* 90–110

James Ernstmeyer, 'Does Strict Territoriality Toll the End of Software Patents Note' (2009) 89(4) *Boston University Law Review* 1267–1304

Robin Feldman and John Newman, 'Copyright at the Bedside: Should We Stop the Spread?' (2012) 16(3) *Stanford Technology Law Review* 623–656

David Ferrance, 'Economic Interests and Jacobsen v. Katzer: Why Open Source Software Deserves Protection under Copyright Law' (2010) 58(4) *Journal of the Copyright Society of the USA* 819

Primavera De Filippi, 'Law of the Cloud: On the Supremacy of the User Interface over Copyright Law' (2013) 2(3) *Internet Policy Review*

Christian O. Fisch et al., 'University patenting: a comparison of 300 leading universities worldwide' (2015) 40(2) *The Journal of Technology Transfer* 318–345

Walter M. Fitch and Emanuel Margoliash, 'Construction of phylogenetic trees' (1967) 155(3760) *Science (New York, N.Y.)* 279–284

Anne Fitzgerald and Christina Cifuentes, 'Accommodating Computer Software to Copyright Doctrine: Defining the Scope of Copyright Protection for Software' (2000) 11(2) *Journal of Law and Information Science* 224–253

Brian Fitzgerald and Nic Suzor, 'Legal Issues for the Use of Free and Open Source Software in Government' (2005) 29(2) *Melbourne University Law Review* 412

Brian Fitzgerald et al., 'Innovation, Software, and Reverse Engineering' (2002) 18(1) *Santa Clara High Technology Law Journal* 121

Richard Fontana, 'Seven Notable Legal Developments In Open Source In 2016' (2017) 8(1) *International Free and Open Source Software Law Review* 59–65

Chiara Franzoni and Giuseppe Scellato, 'Academic Patenting and the Consequences for Scientific Research' (2011) 44(1) *Australian Economic Review* 95–101

Leonard P. Freedman, Iain M. Cockburn and Timothy S. Simcoe, 'The Economics of Reproducibility in Preclinical Research' (2015) 13(6) *PLOS Biol* e1002165

Sascha Friesike et al., 'Opening Science: Towards an Agenda of Open Science in Academia and Industry' (2015) 40(4) *The Journal of Technology Transfer* 581–601

Brett M. Frischmann, 'An Economic Theory of Infrastructure and Commons Management' (2004) 89(4) *Minnesota Law Review* 917–1030

Brett M. Frischmann, 'Two enduring lessons from Elinor Ostrom' (2013) 9(4) *Journal of Institutional Economics* 387–406

Jeanne C. Fromer, 'Patent Disclosure' (2008) 94(2) *Iowa Law Review* 539–606

A. Michael Froomkin, 'Habermas@Discourse.Net: Toward a Critical Theory of Cyberspace' (2002) 116(3) *Harvard Law Review* 749–873

Mark Funk, 'Patent Sharing by US Universities: An Examination of University Joint Patenting' (2013) 22(4) *Economics of Innovation and New Technology* 373–391

Robert T. Furbank and Mark Tester, 'Phenomics – Technologies to Relieve the Phenotyping Bottleneck' (2011) 16(12) *Trends in Plant Science* 635–644

Jeffrey L. Furman, Kyle Jensen and Fiona Murray, 'Governing Knowledge in the Scientific Community: Exploring the Role of Retractions in Biomedicine' (2012) 41(2) *Research Policy* 276–290

Jeffrey L. Furman and Scott Stern, 'Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research' (2011) 101(5) *American Economic Review* 1933–1963

Alberto Galasso, Mark Schankerman and Carlos J. Serrano, 'Trading and Enforcing Patent Rights' (2013) 44(2) *The RAND Journal of Economics* 275–312

Joshua S. Gans, Fiona E. Murray and Scott Stern, 'Contracting Over the Disclosure of Scientific Knowledge: Intellectual Property and Academic Publication' (2017) 46(4) *Research Policy* 820–835

Miguel Garcıa-Sancho, 'A New Insight into Sanger's Development of Sequencing: From Proteins to DNA, 1943–1977' (2010) 43(2) *Journal of the History of Biology* 265–323

Miguel Garcıa-Sancho, 'Genetic Information in the Age of DNA Sequencing' (2015) 50(1) *Information & Culture* 110–142

Miguel Garcıa-Sancho, 'The Proactive Historian: Methodological Opportunities Presented by The New Archives Documenting Genomics' (2016) 55(February) *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 70–82

Paul Gardner et al., 'A Meta-Analysis of Bioinformatics Software Benchmarks Reveals that Publication-Bias Unduly Influences Software Accuracy' (2016) *bioRxiv* 092205

Timothy S. Gardner, Charles R. Cantor and James J. Collins, 'Construction of a Genetic Toggle Switch in *Escherichia coli*' (2000) 403(6767) *Nature* 339–342

Eugene Garfield, 'The History and Meaning of the Journal Impact Factor' (2006) 295(1) *Journal of the American Medical Association* 90–93

Christophe Geiger, Giancarlo Frosio and Oleksandr Bulayenko, 'Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?' (2018) 49(7) *IIC - International Review of Intellectual Property and Competition Law* 814–844

Christophe Geiger, Daniel Gervais and Martin Senftleben, 'The Three-Step Test Revisited: How to Use the Test's Flexibility in National Copyright Law' (2013) 29(3) *American University International Law Review* 581–626

R. Stuart Geiger et al., 'The Types, Roles, and Practices of Documentation in Data Analytics Open Source Software Libraries' (2018) 27(3) *Computer Supported Cooperative Work (CSCW)* 767–802

Robert C. Gentleman et al., 'Bioconductor: Open Software Development for Computational Biology and Bioinformatics' (2004) 5(10) *Genome Biology* R80

Aldo Geuna and Federica Rossi, 'Changes to University IPR Regulations in Europe and the Impact on Academic Patenting' (2011) 40(8) *Research Policy* 1068–1076

Gustavo Ghidini and Emanuela Arezzo, 'Dynamic Competition in Software Development: How Copyrights and Patents, and Their Overlapping, Impact on Derivative Innovation' (2013) 3(4) *Queen Mary Journal of Intellectual Property* 278–295

Shubha Ghosh, 'Are Universities Special' (2016) 49(3) *Akron Law Review* 671–694

Shubha Ghosh, 'Patents and the Regulatory State: Rethinking the Patent Bargain Metaphor after Eldred' (2004) 19(4) *Berkeley Technology Law Journal* 1315

Richard Gilbert and Carl Shapiro, 'Optimal Patent Length and Breadth' (1990) 21(1) *The RAND Journal of Economics* 106

Ben Giles, 'Consideration and the Open Source Agreement' (2002) 49(September) *NSW Society for Computers and the Law*

Jim Giles, 'Software company bans competitive users' (2004) 429(6989) *Nature* 231–231

Jane C. Ginsburg, 'Conflicts of Copyright Ownership between Authors and Owners of Orignial Artworks: An Essay in Comparative and International Private Law' (1992) 17(4) *Columbia-VLA Journal of Law & the Arts* 395–416

Jane C. Ginsburg, 'Four Reasons and a Paradox: The Manifest Superiority of Copyright Over Sui Generis Protection of Computer Software' (1994) 94(8) *Columbia law review* 2559–2572

Donna M. Gitter, 'Resolving the Open Source Paradox in Biotechnology: A Proposal for a Revised open Source Policy for Publicly Funded Genomic Databases' (2006) 43(5) *Houston Law Review* 1475

Rajeev K. Goel and Devrim Göktepe-Hultén, 'What Drives Academic Patentees to Bypass TTOs? Evidence From a Large Public Research Organisation' (2017) *The Journal of Technology Transfer* 1–19

Robert W. Gomulkiewicz, ‘Conditions and Covenants in License Contracts: Tales from a Test of the Artistic License’ (2008) 17(3) *Texas Intellectual Property Law Journal* 335–362

Robert W. Gomulkiewicz, ‘Enforcement of Open Source Software Licenses: The MDY Trio’s Inconvenient Complications’ (2011) 14(1) *Yale Journal of Law and Technology* 106

Andres Guadamuz Gonzalez, ‘Open Science: Open Source Licenses in Scientific Research’ (2005) 7(2) *North Carolina Journal of Law & Technology* 321–366

Jesus M. Gonzalez-Barahona et al., ‘Macro-level Software Evolution: A Case Study of a Large Software Compilation’ (2009) 14(3) *Empirical Software Engineering* 262–285

Patrick R. Goold, ‘Why the U.K. Adaptation Right Is Superior to the U.S. Derivative Work Right’ (2013) 92(4) *Nebraska Law Review* 843–896

Wendy J. Gordon, ‘Fair Use as Market Failure: A Structural and Economic Analysis of the "Betamax" Case and Its Predecessors’ (1982) 82(8) *Columbia Law Review* 1600–1657

Gregory Goth, ‘Bringing Big Data to the Big Tent’ (2015) 58(7) *Communications of the ACM* 17–19

Stuart Graham and Saurabh Vishnubhakat, ‘Of Smart Phone Wars and Software Patents’ (2013) 27(1) *The Journal of Economic Perspectives* 67–85

Emily Grant, Megan Van den Hof and E. Richard Gold, ‘Patent Landscape Analysis: A Methodology in Need of Harmonized Standards of Disclosure’ (2014) 39(4) *World Patent Information* 3–10

Björn A. Grüning et al., ‘Jupyter and Galaxy: Easing Entry Barriers into Complex Data Analyses for Biomedical Researchers’ (2017) 13(5) *PLOS Computational Biology* e1005425

Andres Guadamuz, ‘Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works’ (2017) 2017(2) *Intellectual Property Quarterly* 169–186

Andres Guadamuz and Andrew Rens, 'Comparative Analysis of Copyright Assignment and Licence Formalities for Open Source Contributor Agreements' (2013) 10(2) *SCRIPTed: A Journal of Law, Technology and Society* 207–230

Andrés Guadamuz-González, 'The License/Contract Dichotomy in Open Licenses: a Comparative Analysis' (2009) 30(2) *University of La Verne Law Review* 101

Christi J. Guerrini et al., 'Constraints on Gene Patent Protection Fuel Secrecy Concerns: a Qualitative Study' (2017) 4(3) *Journal of Law and the Biosciences* 542–564

Anna Kaarina Haapanen, 'Free and Open Source Software and the Mystery of Software Patent Licenses Under the GPL' (2015) 7(1) *International Free and Open Source Software Law Review* 19–28

Carolin Haeussler, 'Information-Sharing in Academia and the Industry: A Comparative Study' (2011) 40(1) *Research Policy* 105–122

Carolin Haeussler and Jeannette A. Colyvas, 'Breaking the Ivory Tower: Academic Entrepreneurship in the Life Sciences in UK and Germany' (2011) 40(1) *Research Policy* 41–54

Joel Hagen, 'The Statistical Frame of Mind in Systematic Biology from Quantitative Zoology to Biometry' (2003) 36(2) *Journal of the History of Biology* 353–384

Joel B Hagen, 'The Introduction of Computers into Systematic Research in the United States During the 1960s' (2001) 32(2) *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 291–314

Joel B. Hagen, 'The Origins of Bioinformatics' (2000) 1(3) *Nature Reviews Genetics* 231–236

Thomas Haigh and Mark Priestley, 'Where Code Comes From: Architectures of Automatic Control from Babbage to Algol' (2016) 59(1) *Communications of the ACM* 39–44

Jack S. Hale et al., 'Containers for Portable, Productive, and Performant Scientific Computing' (2017) 19(6) *Computing in Science Engineering* 40–50

Michael Halewood et al., 'Plant Genetic Resources for Food and Agriculture: Opportunities and Challenges Emerging from the Science and Information Technology Revolution' (2018) 217(4) *New Phytologist* 1407–1419

Bronwyn H. Hall and Megan MacGarvie, 'The private value of software patents' (2010) 39(7) *Research Policy* 994–1009

Olof Hallonsten, 'The Third Sector of R&D: Literature Review, Basic Analysis, and Research Agenda' (2017) 35(1) *Prometheus* 21–35

Barbara Hanson, 'Wither Qualitative/Quantitative?: Grounds for Methodological Convergence' (2008) 42(1) *Quality & Quantity* 97–111

Dietmar Harhoff, Frederic M Scherer and Katrin Vopel, 'Citations, Family Size, Opposition and the Value of Patent Rights' (2003) 32(8) *Research Policy* 1343–1363

Mark Harvey and Andrew McMeekin, 'Public or Private Economies of Knowledge: The Economics of Diffusion and Appropriation of Bioinformatics Tools' (2009) 4(1) *International Journal of the Commons* 481–506

Anne-Wil Harzing, Satu Alakangas and David Adams, 'hIa: An Individual Annual h-index to Accommodate Disciplinary and Career Length Differences' (2014) 99(3) *Scientometrics* 811–821

Jerry Hausman, Bronwyn H. Hall and Zvi Griliches, 'Econometric Models for Count Data with an Application to the Patents-R & D Relationship' (1984) 52(4) *Econometrica* 909–938

Bernadine Healy, 'On Gene Patenting' (1992) 327(9) *New England Journal of Medicine* 664–668

Carol A. Heimer, 'Competing Institutions: Law, Medicine, and Family in Neonatal Intensive Care' (1999) 33(1) *Law & Society Review* 17–66

Michael Heller, 'The Tragedy of the Anticommons: A Concise Introduction and Lexicon' (2013) 76(1) *Modern Law Review* 6–25

Michael A. Heller, 'The Tragedy of the Anticommons: Property in the Transition from Marx to Markets' (1997) 111(3) *Harvard Law Review* 621–688

Michael A. Heller and Rebecca S. Eisenberg, 'Can Patents Deter Innovation? The Anticommons in Biomedical Research' (1998) 280(5364) *Science* 698–701

Armijn Hemel and Shane Martin Coughlan, 'Making Sense Of Git In A Legal Context' (2018) 9(1) *International Free and Open Source Software Law Review* 19–33

Rebecca Henderson, Adam B. Jaffe and Manuel Trajtenberg, 'Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965–1988' (1998) 80(1) *The Review of Economics and Statistics* 119–127

Joachim Henkel, 'Selective Revealing in Open Innovation Processes: The Case of Embedded Linux' (2006) 35(7) *Research policy* 953–969

Joachim Henkel and Stephen M. Maurer, 'The Economics of Synthetic Biology' (2007) 3(1) *Molecular Systems Biology* 117

Charlotte Hess and Elinor Ostrom, 'A Framework for Analysing the Microbiological Commons' (2006) 58(188) *International Social Science Journal* 335–349

Jorge. E. Hirsch, 'An Index to Quantify an Individual's Scientific Research Output' (2005) 102(46) *Proceedings of the National Academy of Sciences* 16569–16572

Alexandre Hocquet and Frédéric Wieber, ''Only the Initiates Will Have the Secrets Revealed': Computational Chemists and the Openness of Scientific Software' (2017) 39(4) *IEEE Annals of the History of Computing* 40–58

Thomas Hoeren, 'The Semiconductor Chip Industry – The History, Present and Future of Its IP Law Framework' (2016) 47(7) *IIC - International Review of Intellectual Property and Competition Law* 763–796

Paulien Hogeweg, 'Simulating the Growth of Cellular Forms' (1978) 31(3) *Simulation* 90–96

Paulien Hogeweg, 'The Roots of Bioinformatics in Theoretical Biology' (2011) 7(3) *PLOS Computational Biology* e1002021

Paulien Hogeweg and Ben Hesper, 'Interactive Instruction on Population Interactions' (1978) 8(4) *Computers in Biology and Medicine* 319–327

Timothy R. Holbrook and Mark D. Janis, 'Patent-Eligible Processes: An Audience Perspective' (2014) 17(2) *Vanderbilt Journal of Entertainment and Technology Law* 349–384

Christopher M. Holman, Claes Gustafsson and Andrew W. Torrance, 'Are Engineered Genetic Sequences Copyrightable?: The U.S. Copyright Office Addresses a Matter of First Impression' (2016) 35(3) *Biotechnology Law Report* 103–111

Lee Hood, 'A Personal Journey of Discovery: Developing Technology and Changing Biology' (2008) 1(1) *Annual Review of Analytical Chemistry* 1–43

Leroy Hood et al., 'Systems Biology and New Technologies Enable Predictive and Preventative Medicine' (2004) 306(5696) *Science* 640–643

Julian P Hoppner, 'The GPL Prevails: An Analysis of the First-Ever Court Decision on the Validity and Effectivity of the GPL' (2004) *SCRIPT-ed* 628–635

James Howison and Julia Bullard, 'Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature' (2015) *Journal of the Association for Information Science and Technology*

James Howison et al., 'Understanding the Scientific Software Ecosystem and its Impact: Current and Future Measures' (2015) 24(4) *Research Evaluation* 454–470

Kenneth G. Huang and Fiona E. Murray, 'Entrepreneurial Experiments in Science Policy: Analyzing the Human Genome Project' (2010) 39(5) *Research Policy* 567–582

Mary Huang, 'The Problems of Openness – Effective Regulation of Open Source Software' (2012) 1(1) *International Journal of Technology Policy and Law* 48–68

Wan-Ling Huang, Mary K. Feeney and Eric W. Welch, 'Organizational and Individual Determinants of Patent Production of Academic Scientists and Engineers in the United States' (2011) 38(6) *Science and Public Policy* 463–479

Anton Hughes, 'Avoiding the Software Patent Problem: An Alternative Fix for TRIPS Junkies' (2007) 14(1) *eLaw: Murdoch University Electronic Journal of Law* 100–116

Anton Hughes, 'Bilski v Kappos Case Brief' (2010) 20(1) *Journal of Law, Information and Science* 206–213

Tommi Inkinen and Katri Suorsa, 'Intermediaries in Regional Innovation Systems: High-Technology Enterprise Survey from Northern Finland' (2010) 18(2) *European Planning Studies* 169–187

'Integrating Ethics and Science in the International HapMap Project' , (2004) 5(6) *Nature reviews. Genetics* 467–475

International HapMap Consortium, 'The International HapMap Project' (2003) 426(6968) *Nature* 789–796

Thomas Jaeger, 'Reset and Go: The Unitary Patent System Post-Brexit' (2017) 48(3) *IIC - International Review of Intellectual Property and Competition Law* 254–285

Adam B. Jaffe and Manuel Trajtenberg, 'Flows of Knowledge from Universities and Federal Laboratories: Modeling the Flow of Patent Citations over Time and Across Institutional and Geographic Boundaries' (1996) 93(23) *Proceedings of the National Academy of Sciences* 12671–12677

Harrie Jansen, 'The Logic of Qualitative Survey Research and its Position in the Field of Social Research Methods' (2010) 11(2) *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*

Dennis M. Jennings et al., 'Computer Networking for Scientists' (1986) 231(4741) *Science* 943–950

Jenny Fry, Ralph Schroeder and Matthijs den Besten, 'Open Science in e-Science: Contingency or Policy?' (2009) 65(1) *Journal of Documentation* 6–32

Omar Johnny, Marc Miller and Mark Webbink, 'Copyright in Open Source Software - Understanding the Boundaries' (2010) 2(1) *International Free and Open Source Software Law Review* 13–38

Ann Johnson, 'Modeling Molecules: Computational Nanotechnology as a Knowledge Community' (2009) 17(2) *Perspectives on Science* 144–173

Jeroen P. J. de Jong, Tarmo Kalvet and Wim Vanhaverbeke, 'Exploring a Theoretical Framework to Structure the Public Policy Implications of Open Innovation' (2010) 22(8) *Technology Analysis & Strategic Management* 877–896

Justin Joque, ‘The Invention of the Object: Object Orientation and the Philosophical Development of Programming Languages of Programming Languages’ (2016) 29(4) *Philosophy & Technology* 335–356

Ali K. Yetisen and Lisa R. Volpatti, ‘Patent Protection and Licensing in Microfluidics’ (2014) 14(13) *Lab on a Chip* 2217–2225

Linda J. Kahl and Drew Endy, ‘A survey of enabling technologies in synthetic biology’ (2013) 7(1) *Journal of Biological Engineering* 13

Pascal Kahlem and Ewan Birney, ‘Dry Work in a Wet World: Computation in Systems Biology’ (2006) 2(1) *Molecular Systems Biology* 40

Dennis S. Karjala, ‘Lessons from the Computer Software Protection Debate in Japan’ (1984) *Arizona State Law Journal* 53–82

Dennis S. Karjala, ‘Protecting Innovation in Computer Software, Biotechnology, and Nanotechnology’ (2011) 16(1) *Virginia Journal of Law & Technology* 42–65

Ariel Katz, ‘A Network Effects Perspective on Software Piracy’ (2005) 55(2) *University of Toronto Law Journal* 155–216

Ariel Katz, ‘Substitutions and Schumpeterian Effects over the Life Cycle of Copyrighted Works’ (2008) 49(2) *Jurimetrics* 113–154

Martin Kenney and Donald Patton, ‘Reconsidering the Bayh-Dole Act and the Current University Invention Ownership Model’ (2009) 38(9) *Research Policy* 1407–1422

Wolfgang Kerber and Heike Schweitzer, ‘Interoperability in the Digital Economy’ (2017) 8(1) *Journal of Intellectual Property, Information Technology and E-Commerce Law* 39–58

Jack Kloppenburg, ‘Impeding Dispossession, Enabling Repossession: Biological Open Source and the Recovery of Seed Sovereignty’ (2010) 10(3) *Journal of Agrarian Change* 367–388

Samuel Kortum and Josh Lerner, ‘What is Behind the Recent Surge in Patenting?’ (1999) 28(1) *Research Policy* 1–22

Georg von Krogh, Sebastian Spaeth and Karim R Lakhani, 'Community, Joining, and Specialization in Open Source Software Innovation: a Case Study' (2003) 32(7) *Research Policy* 1217–1241

Katharine Ku and James Henderson, 'The MTA—Rip it Up and Start Again?' (2007) 25(7) *Nature Biotechnology* 721–721

Paul B. de Laat, 'Copyright or copyleft?: An analysis of property regimes for software development' (2005) 34(10) *Research Policy* 1511–1532

Eric S. Lander et al., 'Initial Sequencing and Analysis of the Human Genome' (2001) 409(6822) *Nature* 860–921

William M. Landes and Richard A. Posner, 'An Economic Analysis of Copyright Law' (1989) 18(2) *Journal of Legal Studies* 325–364

Christoph Laub, 'Software Patenting: Legal Standards in Europe and the US in view of Strategic Limitations of the IP Systems' (2006) 9(3) *The Journal of World Intellectual Property* 344–372

Charles Lawson and Michelle Rourke, 'Open Access DNA, RNA and Amino Acid Sequences: The Consequences and Solutions for the International Regulation of Access and Benefit Sharing' (2016) 24(3) *Journal of Law and Medicine* 96–118

Joshua Lederberg, 'Digital Communications and the Conduct of Science: The New Literacy' (1978) 66(11) *Proceedings of the IEEE* 1314–1319

David Lehr and Paul Ohm, 'Playing with the Data: What Legal Scholars Should Learn about Machine Learning' (2017) 51(2) *U.C. Davis Law Review* 653–718

Zhen Lei, Rakhi Juneja and Brian D. Wright, 'Patents versus Patenting: Implications of Intellectual Property Protection for Biological Research' (2009) 27(1) *Nature Biotechnology* 36–40

Jeremy Leipzig, 'A Review of Bioinformatic Pipeline Frameworks' (2017) 18(3) *Briefings in Bioinformatics* 530–536

Philip Leith, 'Patenting Programs as Machines' (2007) 4(2) *SCRIPTed: A Journal of Law, Technology and Society* 214–226

Doug Lichtman & Mark A. Lemley, 'Rethinking Patent Law's Presumption of Validity' (2007) 60(1) *Stanford Law Review* 45

Mark A. Lemley, 'Convergence in the Law of Software Copyright' (1995) 10(1) *High Technology Law Journal* 1–34

Josh Lerner and Jean Tirole, 'The Scope of Open Source Licensing' (2005) 21(1) *The Journal of Law, Economics, and Organization* 20–56

Joshua Lerner, 'The Importance of Patent Scope: An Empirical Analysis' (1994) 25(2) *The RAND Journal of Economics* 319–333

Pierre N. Leval, 'Toward a Fair Use Standard' (1989) 103(5) *Harvard Law Review* 1105–1136

David Levi-Faur, 'Regulatory Capitalism: The Dynamics of Change beyond Telecoms and Electricity' (2006) 19(3) *Governance* 497–525

Nadine Levin and Sabina Leonelli, 'How Does One "Open" Science? Questions of Value in Biological Research' (2016) *Science, Technology, & Human Values* 0162243916672071

Elliott C. Levinthal et al., 'When Computers Talk To Computers' (1975) 17(12) *Industrial Research* 35–42

Jamie Lewis and Andrew Bartlett, 'Inscribing a Discipline: Tensions in the Field of Bioinformatics' (2013) 32(3) *New Genetics and Society* 243–263

Joanna Lewis et al., 'Where Next for the Reproducibility Agenda in Computational Biology?' (2016) 10(1) *BMC Systems Biology* 52

Maxwell W. Libbrecht and William Stafford Noble, 'Machine learning applications in genetics and genomics' (2015) 16(6) *Nature Reviews Genetics* 321–332

Evan S. Lieberman, 'Nested Analysis as a Mixed-Method Strategy for Comparative Research' (2005) 99(3) *American Political Science Review* 435–452

Aaron M. Lien, Edella Schlager and Ashly Lona, 'Using institutional grammar to improve understanding of the form and function of payment for ecosystem services programs' (2018) 31() *Ecosystem Services* 21–31

Francesco Lissoni, Fabio Montobbio and Lorenzo Zirulia, 'Inventorship and Authorship as Attribution Rights: An Enquiry Into the Economics of Scientific Credit' (2013) 95(11) *Journal of Economic Behavior & Organization* 49–69

Francesco Lissoni et al., 'Academic Patenting in Europe: New Evidence from the KEINS Database' (2008) 17(2) *Research Evaluation* 87–102

Michelle Marie Lockhart, Zaheer-Ud-Din Babar and Sanjay Garg, 'Drug development and research in New Zealand: policies affecting the industry' (2012) 73(1) *Drug Development Research* 1–10

Clarisa Long, 'Patent Signals' (2002) 69(2) *The University of Chicago Law Review* 625–679

Brian J. Love, 'Do University Patents Pay Off-Evidence from a Survey of University Inventors in Computer Science and Electrical Engineering' (2013) 16(2) *Yale Journal of Law and Technology* 285

Nicola Lucchi, 'Understanding Genetic Information as a Commons: From Bioprospecting to Personalized Medicine' (2013) 7(2) *International Journal of the Commons* 313–338

Nicholas M. Luscombe, Dov Greenbaum and Mark Gerstein, 'What is Bioinformatics? A Proposed Definition and Overview of the Field' (2001) 40(4) *Methods of information in medicine* 346–358

Robert D. Macredie and Kabiru Mijinyawa, 'A Theory-Grounded Framework of Open Source Software Adoption in SMEs' (2011) 20(2) *European Journal of Information Systems* 237–250

Michael J. Madison, Brett M. Frischmann and Katherine J. Strandburg, 'Constructing Commons in the Cultural Environment' (2009) 95(4) *Cornell Law Review* 657–710

Tom Magerman, Bart Van Looy and Koenraad Debackere, 'Does Involvement in Patenting Jeopardize One's Academic Footprint? An Analysis of Patent-Paper Pairs in Biotechnology' (2015) 44(9) *Research Policy* 1702–1713

Peter B. Maggs, 'License Contracts, Free Software and Creative Commons in the United States' (2014) 62(suppl_1) *The American Journal of Comparative Law* 407–423

Neha Mago, Nishad Deshpande and Rajkumar R. Hirwani, 'A Landscape of Bioinformatics Patents - Garnering of IPR in the Field of Bioinformatics' (2017) 51(4) *World Patent Information* 66–78

Teri A. Manolio et al., 'New Models of Collaboration in Genome-Wide Association Studies: the Genetic Association Information Network' (2007) 39(9) *Nature Genetics* 1045–1051

Elaine R. Mardis, 'The Impact of Next-Generation Sequencing Technology on Genetics' (2008) 24(3) *Trends in Genetics* 133–141

Bryan Marshall et al., 'Does Sample Size Matter in Qualitative Research?: A Review of Qualitative Interviews in IS Research' (2013) 54(1) *Journal of Computer Information Systems* 11–22

Keith Maskus, 'The New Globalisation of Intellectual Property Rights: What's New This Time?' (2014) 54(3) *Australian Economic History Review* 262–284

Mark Mason, 'Sample Size and Saturation in PhD Studies Using Qualitative Interviews' (2010) 11(3) *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*

Benjamin J. McEniery, 'An Empirical Study of Business Method Patent Applications Filed in Australia 2000-2009' (2012) -(89) *Intellectual Property Forum*

Michael D. McGinnis, 'An Introduction to IAD and the Language of the Ostrom Workshop: A Simple Guide to a Complex Framework' (2011) 39(1) *Policy Studies Journal* 169–183

Heather J. Meeker, 'Open Source and the Age of Enforcement' (2012) 4(2) *Hastings Science & Technology Law Journal* 267–290

Heather Meeker and Stephanie Petit, 'The New Foundations of Open Source' (2016) 33(3) *Santa Clara High Technology Law Journal* 415–426

Rebeca Méndez-Durón and Clara E. Garcıa, 'Returns from social capital in open source software networks' (2009) 19(2) *Journal of Evolutionary Economics* 277–295

Peter S. Menell, 'API Copyrightability Bleak House: Unraveling and Repairing the Oracle v. Google Jurisdictional Mess' (2016) 31(3) *Berkeley Technology Law Journal* 1515–1596

Robert P. Merges and Michael Mattioli, 'Measuring the Costs and Benefits of Patent Pools' (2017) 78(2) *Ohio State Law Journal* 281–348

Aexel Metzger, 'Internationalisation of FOSS Contributory Copyright Assignments and Licenses: Jurisdiction-Specific or "Unported"?' (2013) 10(2) *SCRIPTed* 177

Martin Michlmayr, 'Community Management in Open Source Projects' (2009) X(3) *The European Journal for the Informatics Professional, X (3)* 22–26

Arthur R. Miller, 'Copyright Protection for Computer Programs, Databases, and Computer-Generated Works: Is Anything New Since CONTU' (1992) 106(5) *Harvard Law Review* 977–1073

Timo Minssen, Berthold Rutz and Esther van Zimmeren, 'Synthetic Biology and Intellectual Property Rights: Six Recommendations' (2015) 10(2) *Biotechnology Journal* 236–241

Timo Minssen and Robert M. Schwartz, 'US Patent Eligibility in the Wake of Bilski v Kappos: "Business as Usual" in an Age of New Technologies?' (2011) 30(1) *Biotechnology Law Report* 3–56

Amrita Mishra and Tania Bubela, 'Legal Agreements and the Governance of Research Commons: Lessons from Materials Sharing in Mouse Genomics' (2014) 18(4) *OMICS: A Journal of Integrative Biology* 254–273

Amrita Mishra, Paul N. Schofield and Tania M. Bubela, 'Sustaining Large-Scale Infrastructure to Promote Pre-Competitive Biomedical Research: Lessons from Mouse Genomics' (2016) 33(2) *New Biotechnology* 280–294

Audris Mockus, Roy T. Fielding and James D. Herbsleb, 'Two Case Studies of Open Source Software Development: Apache and Mozilla' (2002) 11(3) *ACM Trans. Softw. Eng. Methodol.* 309–346

Ann L. Monotti, 'The Australian Experimental Use Exemption: A Current Overview' (2009) 12(5) *The Journal of World Intellectual Property* 422–445

Ann L. Monotti, 'The Scope of Manner of Manufacture under the Patents Act 1990 (CTH) after Grant v Commissioner of Patents Comment' (2006) 34(3) *Federal Law Review* 460–478

Andrew Morin, Jennifer Urban and Piotr Sliz, 'A Quick Guide to Software Licensing for the Scientist-Programmer' (2012) 8(7) *PLoS Comput Biol* e1002598

Sarah L. Moritz and Andrew F. Christie, 'Second-Tier Patent Systems: the Australian Experience' (2006) 28(4) *European Intellectual Property Review* 230

Kelsey Martin Mott, 'The Concept of Small Patent in European Legal Systems and Equivalent Protection under United States Law' (1963) 49(2) *Virginia Law Review* 232–261

David C. Mowery and Bhaven N. Sampat, 'The Bayh-Dole Act of 1980 and University–Industry Technology Transfer: A Model for Other OECD Governments?' (2004) 30(1-2) *The Journal of Technology Transfer* 115–127

Fiona Murray, 'Innovation as Co-Evolution of Scientific and Technological Networks: Exploring Tissue Engineering' (2002) 31(8) *Research Policy* 1389–1403

Fiona Murray and Scott Stern, 'Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge?: An Empirical Test of the Anti-Commons Hypothesis' (2007) 63(4) *Journal of Economic Behavior & Organization* 648–687

Vivek Navale and Philip E. Bourne, 'Cloud Computing Applications for Biomedical Science: A Perspective' (2018) 14(6) *PLOS Computational Biology* e1006144

Pieter B. T. Neerincx and Jack A. M. Leunissen, 'Evolution of Web Services in Bioinformatics' (2005) 6(2) *Briefings in Bioinformatics* 178–188

Andrew J. Nelson, 'From the Ivory Tower to the Startup Garage: Organizational Context and Commercialization Processes' (2014) 43(7) *Research Policy* 1144–1156

Victoria Nemiah, 'License and Registration, Please: Using Copyright Conditions to Protect Free/Open Source Software' (2013) 3(2) *New York University Journal of Intellectual Property and Entertainment Law* 358

Marc Nerlove, 'Programming Languages: A Short History for Economists' (2004) 29(1-3) *Journal of Economic & Social Measurement* 189–203

Harvey B. Newman, Mark H. Ellisman and John A. Orcutt, 'Data-Intensive E-Science Frontier Research' (2003) 46(11) *Communications of the ACM* 68–77

Dianne Nicol, 'Cooperative Intellectual Property in Biotechnology' (2007) 4(1) *SCRIPT-ed* 136–151

Dianne Nicol, 'Public Trust, Intellectual Property and Human Genetic Databanks: the Need to Take Benefit Sharing Seriously' (2006) 3(3) *Journal of International Biotechnology Law* 89–103

Dianne Nicol and Jane Nielsen, 'Australian Medical Biotechnology: Navigating a Complex Patent Landscape' (2005) 27(9) *European Intellectual Property Review* 313–318

Jane Nielsen and Dianne Nicol, 'Whither Patent Use without Authorization in Australia' (2008) 36(3) *Federal Law Review* 333–364

Michael Noel and Mark Schankerman, 'Strategic Patenting and Software Innovation' (2013) 61(3) *The Journal of Industrial Economics* 481–520

Joseph November, 'Removing the Center from Computing: Biology's New Mode of Digital Knowledge Production' (2011) 34(2) *Berichte zur Wissenschaftsgeschichte* 156–173

Aisling O'Driscoll, Jurate Daugelaite and Roy D. Sleator, "Big Data', Hadoop and Cloud Computing in Genomics' (2013) 46(5) *Journal of Biomedical Informatics* 774–781

Ruth L. Okediji, 'Toward an International Fair Use Doctrine' (2000) 39(1) *Columbia Journal of Transnational Law* 75–176

Ville Oksanen and Mikko Välimäki, 'Free Software and Copyright Enforcement: A Tool for Global Copyright Policy?' (2006) 18(4) *Knowledge, Technology & Policy* 101–112

Paul Oldham and Anthony Mark Cutter, 'Mapping Global Status and Trends in Patent Activity for Biological and Genetic Material' (2006) 2(2) *Genomics, Society and Policy* 62

Paul Oldham, Stephen Hall and Geoff Burton, 'Synthetic Biology: Mapping the Scientific Landscape' (2012) 7(4) *PLoS ONE* e34368

Siobhán O'Mahony, 'Guarding the Commons: How Community Managed Software Projects Protect their Work' (2003) 32(7) *Research Policy* 1179–1198

Trent Ostler and Michael Gollin, 'Which Types of Bioinformatics Inventions are Eligible for Patent Protection?' (2015) 21(2) *Journal of Commercial Biotechnology* 76–82

Elinor Ostrom, 'Background on the Institutional Analysis and Development Framework' (2011) 39(1) *Policy Studies Journal* 7–27

Elinor Ostrom, 'Beyond Markets and States: Polycentric Governance of Complex Economic Systems' (2010) 100(3) *The American Economic Review* 641–672

Lisa Larrimore Ouellette, 'Patent Experimentalism' (2015) 101(1) *Virginia Law Review* 65–128

Lisa Larrimore Ouellette, 'Who Reads Patents?' (2017) 35(5) *Nature Biotechnology* 421–424

Joey Paquet and Serguei A. Mokhov, 'Comparative Studies of Programming Languages; Course Lecture Notes' (2010) *arXiv:1007.2123 [cs]*

Gideon Parchomovsky and Michael Mattioli, 'Partial Patents' (2011) 111(2) *Columbia Law Review* 207

Gideon Parchomovsky and R. Polk Wagner, 'Patent Portfolios' (2005) 154(1) *University of Pennsylvania Law Review* 1–77

Gideon Parchomovsky and Philip J. Weiser, 'Beyond Fair Use' (2010) 96(2) *Cornell Law Review* 91–138

Hyun-Seok Park, 'Preliminary Study of Bioinformatics Patents and Their Classifications Registered in the KIPRIS Database' (2012) 10(4) *Genomics & Informatics* 271–274

Jinseok Park, 'Evolution of Industry Knowledge in the Public Domain: Prior Art Searching for Software Patents' (2005) 2(1) *SCRIPTed: A Journal of Law, Technology and Society* 47–70

United States Patent and Trademark Office, 'Interim Guidance for Determining Subject Matter Eligibility for Process Claims in View of Bilski v. Kappos' (27 July 2010) 75(43922) *F.R.*

Markus Perkmann and Henri Schildt, 'Open Data Partnerships between Firms and Universities: The Role of Boundary Organizations' (2015) 44(5) *Research Policy* 1133–1143

Justine Pila, 'Dispute Over the Meaning of 'Invention' in Article 52(2) EPC: The patentability of computer-implemented inventions in Europe' (2005) 36(2) *IIC: International Review of Industrial Property & Copyright Law* 171

Justine Pila, ‘Software Parents, Separation of Powers, and Failed Syllogisms: A Cornucopia from the Enlarged Board of Appeal of the European Office’ (2011) 70(2) *Cambridge Law Journal* 203–228

Heather A. Piwowar and Todd J. Vision, ‘Data Reuse and the Open Data Citation Advantage’ (2013) [1] *PeerJ* e175

Hans E. Plesser, ‘Reproducibility vs. Replicability: A Brief History of a Confused Terminology’ (2018) *Frontiers in Neuroinformatics*

Mihai Pop and Steven L. Salzberg, ‘Bioinformatics Challenges of New Sequencing Technology’ (2008) 24(3) *Trends in Genetics* 142–149

David Popp, ‘They Don’t Invent Them Like They Used to: An Examination of Energy Patent Citations Over Time’ (2006) 15(8) *Economics of Innovation and New Technology* 753–776

Walter W. Powell et al., ‘Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences’ (2005) 110(4) *American Journal of Sociology* 1132–1205

John S. Quarterman, Abraham Silberschatz and James L. Peterson, ‘4.2BSD and 4.3BSD As Examples of the UNIX System’ (1985) 17(4) *ACM Computing Surveys* 379–418

Mark A. Ragan, Tim Littlejohn and Bruce Ross, ‘Genome-Scale Computational Biology and Bioinformatics in Australia’ (2008) 4(8) *PLOS Comput Biol* e1000068

Arti K. Rai, John R. Allison and Bhaven N. Sampat, ‘University Software Ownership and Litigation: A First Examination’ (2009) 87(5) *North Carolina Law Review* 1519–1570

Arti Rai and James Boyle, ‘Synthetic Biology: Caught between Property Rights, the Public Domain, and the Commons’ (2007) 5(3) *PLoS Biology*

Karthik Ram, ‘Git can facilitate greater reproducibility and increased transparency in science’ (2013) 8(1) *Source Code for Biology and Medicine* 7

Benjamin Raue, 'Free Flow of Data? The Friction Between the Commission's European Data Economy Initiative and the Proposed Directive on Copyright in the Digital Single Market' (2018) 49(4) *IIC - International Review of Intellectual Property and Competition Law* 379–383

Jerome H. Reichman, 'Legal Hybrids between the Patent and Copyright Paradigms' (1994) 94(8) *Columbia Law Review* 2432

Jerome H. H. Reichman and Paul F. Uhlir, 'A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment' (2003) 66(1/2) *Law and Contemporary Problems* 315–462

Jerome H. Reichman and Ruth L. Okediji, 'When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale' (2012) 96(4) *Minnesota Law Review* 1362–1480

John. R. Rice and Ronald. F. Boisvert, 'From Scientific Software Libraries to Problem-Solving Environments' (1996) 3(3) *IEEE Computational Science and Engineering* 44–53

Megan Richardson, 'Sui Generis Intellectual Property Law Reform: Issues for Australia' (2001) 32(1) *Victoria University of Wellington Law Review* 19–46

Matthew Rimmer, 'The Freedom to Tinker: Patent Law and Experimental Use' (2005) 15(2) *Expert Opinion on Therapeutic Patents* 167–200

Dennis M. Ritchie, 'Reflections on Software Research' (2012) 17(8) *Resonance* 810–816

L. Roberts, 'NIH gene patents, round two' (1992) 255(5047) *Science* 912–913

Leslie Roberts, 'Controversial From the Start' (2001) 291(5507) *Science* 1182–1188

Leslie Roberts, 'The Perils of Involving Congress in a 'Catfight'' (1992) 257(5067) *Science* 156–157

Ann R. R. Robertson et al., 'Tightrope walking towards maximising secondary uses of digitised health data: a qualitative study.' (2016) 23(3) *Journal of Innovation in Health Informatics* 591–599

Peter W. Rose et al., 'The RCSB Protein Data Bank: Integrative View of Protein, Gene and 3D Structural Information' (2017) 45(D1) *Nucleic Acids Research* D271–D281

Simone A. Rose, 'Semiconductor Chips, Genes, and Stem Cells: New Wine for New Bottles?' (2012) 38(1) *American Journal of Law & Medicine* 113

Arnon Rosenthal et al., 'Cloud Computing: A New Business Paradigm for Biomedical Information Sharing' (2010) 43(2) *Journal of Biomedical Informatics* 342–353

Somak Roy et al., 'Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists' (2018) 20(1) *The Journal of Molecular Diagnostics* 4–27

Andrew L. Russell and Valérie Schafer, 'In the Shadow of ARPANET and Internet: Louis Pouzin and the Cyclades Network in the 1970s' (2014) 55(4) *Technology and Culture* 880–907

Matthew Sag, 'Predicting Fair Use' (2012) 73(1) *Ohio State Law Journal* 47–92

Matthew Sag, 'The Prehistory of Fair Use' (2010) 76(4) *Brooklyn Law Review* 1371–1412

Negin Salimi, Rudi Bekkers and Koen Frenken, 'Does Working with Industry Come at a Price? A Study of Doctoral Candidates' Performance in Collaborative vs. Non-Collaborative Ph.D. Projects' (2015) 41-42(4) *Technovation* 51–61

Pamela Samuelson, 'Benson Revisited: The Case Against Patent Protection For Algorithms and Other Computer Program-Related Inventions' (1990) 39(4) *Emory Law Journal* 1025–1154

Pamela Samuelson, 'Copyrightability of Java APIs Revisited' (2015) 58(3) *Communications of the ACM* 22–24

Pamela Samuelson, 'Evolving Conceptions of Copyright Subject Manner' (2016) 78(1) *University of Pittsburgh Law Review* 17–94

Pamela Samuelson, 'Functionality and Expression in Computer Programs: Refining the Tests for Software Copyright Infringement' (2016) 31(2) *Berkeley Technology Law Journal* 1215–1300

Pamela Samuelson, 'IBM's Pragmatic Embrace of Open Source' (2006) 49(10) *Communications of the ACM* 21–25

Pamela Samuelson, 'Oracle V. Google: Are APIs Copyrightable?' (2012) 55(11) *Communications of the ACM* 25–27

Pamela Samuelson, 'Should Economics Play a Role in Copyright Law and Policy' (2003) 1(1-2) *University of Ottawa Law & Technology Journal* 1–22

Pamela Samuelson, 'The Uneasy Case for Software Copyrights Revisited' (2010) 79(6) *George Washington Law Review* 1746

Pamela Samuelson and Robert J. Glushko, 'Comparing the Views of Lawyers and User Interface Designers on the Software Copyright Look and Feel Lawsuits' (1989) 30(1) *Jurimetrics Journal* 121–140

Pamela Samuelson and Suzanne Scotchmer, 'The Law and Economics of Reverse Engineering' (2001) 111() *Yale Law Journal* 1575–1664

Pamela Samuelson et al., 'A Manifesto concerning the Legal Protection of Computer Programs' (1994) 94(8) *Columbia Law Review* 2308–2431

Madelyn Sanfilippo, Brett Frischmann and Katherine Standburg, 'Privacy as Commons: Case Evaluation Through the Governing Knowledge Commons Framework' (2018) 8() *Journal of Information Policy* 116–166

Frederick Sanger, 'Sequences, Sequences, and Sequences' (1988) 57(1) *Annual Review of Biochemistry* 1–29

Frederick Sanger, 'The Free Amino Groups of Insulin' (1945) 39(5) *Biochemical Journal* 507–515

Frederick Sanger, 'The Terminal Peptides of Insulin' (1949) 45(5) *Biochemical Journal* 563–574

Robert M. Sauer, 'Why Develop Open-Source Software? The Role of Non-Pecuniary Benefits, Monetary Rewards, and Open-Source Licence Type' (2007) 23(4) *Oxford Review of Economic Policy* 605–619

Tempest A. van Schaik et al., 'The need to redefine genomic data sharing: A focus on data accessibility' (2014) 3(4) *Applied & Translational Genomics* 100–104

James Scheibner and Dianne Nicol, 'Do Software Patents Inhibit Open Source Licensing in Australia?' (2015) 25(4) *Australian Intellectual Property Journal* 198

Martin P. Schmidt, 'Patent Strategies in the Process-Related Industries: Outline of the Problems' (2013) 43(3) *R&D Management* 242–251

Anja Schoen, Bruno van Pottelsberghe de la Potterie and Joachim Henkel, 'Governance Typology of Universities' Technology Transfer Processes' (2014) 39(3) *The Journal of Technology Transfer* 435–453

Michael Schwarz and Yuri Takhteyev, 'Half a Century of Public Software Institutions: Open Source as a Solution to Hold-Up Problem' (2010) 12(4) *Journal of Public Economic Theory* 609–639

Charles M. Schweik, 'An Institutional Analysis Approach to Studying Libre Software Commons' (2005) 6(3) *Upgrade: The European Journal for the Informatics Professional* 17–27

Charles M. Schweik and Meelis Kitsing, 'Applying Elinor Ostrom's Rule Classification Framework to the Analysis of Open Source Software Commons' (2010) 2(1) *Transnational Corporations Review* 13–26

Charles Schweik and Robert English, 'Preliminary Steps Toward a General Theory of Internet-Based Collective-Action in Digital Information Commons: Findings From a Study of Open Source Software Projects' (2013) 7(2) *International Journal of the Commons* 234–254

Charles Schweik, Tom Evans and J. Morgan Grove, 'Open Source and Open Content: a Framework for Global Collaboration in Social-Ecological Research' (2005) 10(1) *Ecology and Society*

Suzanne Scotchmer, 'Standing on the Shoulders of Giants: Cumulative Research and the Patent Law' (1991) 5(1) *The Journal of Economic Perspectives* 29–41

David B. Searls, 'The Roots of Bioinformatics' (2010) 6(6) *PLOS Computational Biology* e1000809

Commonwealth Secretariat, 'The Role of National Intellectual Property Laws in Promoting Innovation, Scientific and Technological Development' (2017) 43(3-4) *Commonwealth Law Bulletin* 471–488

Kenneth C. Shadlen, Andrew Schrank and Marcus J. Kurtz, 'The Political Economy of Intellectual Property Protection: The Case of Software' (2005) 49(1) *International Studies Quarterly* 45–71

Jay Shendure and Hanlee Ji, 'Next-Generation DNA Sequencing' (2008) 26(10) *Nature Biotechnology* 1135–1145

Jacob S. Sherkow, 'Patent Law's Reproducibility Paradox' (2016) 66(4) *Duke Law Journal*

Saba Siddiki, Xavier Basurto and Christopher M. Weible, 'Using the Institutional Grammar Tool to Understand Regulatory Compliance: The Case of Colorado Aquaculture' (2012) 6(2) *Regulation & Governance* 167–188

Donald S Siegel, David Waldman and Albert Link, 'Assessing the Impact of Organizational Practices on the Relative Productivity of University Technology Transfer Offices: an Exploratory Study' (2003) 32(1) *Research Policy* 27–48

Alexandra Sims, 'The Case for Fair Use in New Zealand' (2016) 24(2) *International Journal of Law and Information Technology* 176–202

Lillian L. Siu et al., 'Facilitating a Culture of Responsible and Effective Sharing of Cancer Genome Data' (2016) 22(5) *Nature medicine* 464–471

Benjamin Six et al., 'Trust and social capital in the design and evolution of institutions for collective action' (2015) 9(1) *International Journal of the Commons*

Dennis H. Smith et al., 'BIONET: National Computer Resource for Molecular Biology' (1986) 14(1) *Nucleic Acids Research* 17–20

Lloyd M. Smith et al., 'Fluorescence Detection in Automated DNA Sequence Analysis' (1986) 321(6071) *Nature* 674–679

Anthony D. So et al., 'Is Bayh-Dole Good for Developing Countries? Lessons from the US Experience' (2008) 6(10) *PLOS Biol* e262

Johan Söderberg, ‘Free Space Optics in the Czech Wireless Community: Shedding Some Light on the Role of Normativity for User-Initiated Innovations’ (2011) 36(4) *Science, Technology, & Human Values* 423–450

Sylvia Song, ‘Software patents - a Comparison of the UK/European and Australian Laws on Patentability’ (2015) [102] *Intellectual Property Forum: journal of the Intellectual and Industrial Property Society of Australia and New Zealand* 52

Sören Sonnenburg et al., ‘The Need for Open Source Software in Machine Learning’ (2007) 8(Oct) *Journal of Machine Learning Research* 2443–2466

Jason E. Stajich and Hilmar Lapp, ‘Open Source Tools and Toolkits for Bioinformatics: Significance, and Where Are We?’ (2006) 7(3) *Briefings in Bioinformatics* 287–296

Lincoln D. Stein, ‘The Case for Cloud Computing in Genome Informatics’ (2010) 11(5) *Genome Biology* 207

Sigrid Sterckx and Julian Cockbain, ‘The Patentability of Computer Programs in Europe: An Improved Interpretation of Articles 52(2) and (3) of the European Patent Convention’ (2010) 13(3) *The Journal of World Intellectual Property* 366–402

Hallam Stevens, ‘Networking Biology: The Origins of Sequence-Sharing Practices in Genomics’ (2015) 56(4) *Technology and Culture* 839–867

Hallam Stevens, ‘The Politics of Sequence: Data Sharing and the Open Source Software Movement’ (2015) 50(4) *Information & Culture: A Journal of History* 465–503

Victoria Stodden, ‘Open Science: Policy Implications for the Evolving Phenomenon of User-Led Scientific Innovation’ (2010) 9(1) *Journal of Science Communication* 1–8

Victoria Stodden, Peixuan Guo and Zhaokun Ma, ‘Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals’ (2013) 8(6) *PLOS ONE* e67111

Victoria Stodden et al., ‘Enhancing Reproducibility for Computational Methods’ (2016) 354(6317) *Science* 1240–1241

Klaas-Jan Stol and Brian Fitzgerald, 'Theory-Oriented Software Engineering' (2015) 101() *Science of Computer Programming* 79–98

Anita Stork, 'The Use of Arbitration in Copyright Disputes: IBM v. Fujitsu' (1988) 3(2) *Berkeley Technology Law Journal* 241

Katherine Strandburg, 'Patent Fair Use 2.0' (2011) 1(2) *UC Irvine Law Review* 265

Katherine J. Strandburg, 'What Does the Public Get - Experimental Use and the Patent Bargain' (2004) [1] *Wisconsin Law Review* 81–156

Katherine J. Strandburg, 'What if There Were a Business Method Use Exemption to Patent Infringement' (2008) *Michigan State Law Review* 245–278

Bruno J. Strasser, 'Collecting, Comparing, and Computing Sequences: The Making of Margaret O. Dayhoff's Atlas of Protein Sequence and Structure, 1954–1965' (2009) 43(4) *Journal of the History of Biology* 623–660

Bruno J. Strasser, 'Genetics. GenBank–Natural history in the 21st Century?' (2008) 322(5901) *Science (New York, N.Y.)* 537–538

Bruno J. Strasser, 'The Experimenter's Museum: GenBank, Natural History, and the Moral Economies of Biomedicine' (2011) 102(1) *Isis* 60–96

Edna Suárez-Dıaz, 'Making Room for New Faces: Evolution, Genomics and the Growth of Bioinformatics' (2010) 32(1) *History and Philosophy of the Life Sciences* 65–89

Edna Suárez-Dıaz and Victor H. Anaya-Muñoz, 'History, Objectivity, and the Construction of Molecular Phylogenies' (2008) 39(4) *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 451–468

John Swinson, 'Copyright or Patent or Both: An Algorithmic Approach to Computer Software Protection' (1991) 5(1) *Harvard Journal of Law & Technology* 145–214

Andrew S. Tanenbaum, 'Lessons Learned from 30 Years of MINIX' (2016) 59(3) *Communications of the ACM* 70–78

Y. Tarui, Y. Hayashi and K. Nagai, 'Electrically Reprogrammable Nonvolatile Semiconductor Memory' (1972) 7(5) *IEEE Journal of Solid-State Circuits* 369–375

Neil C. Thompson, Arvids A. Ziedonis and David C. Mowery, 'University Licensing and the Flow of Scientific Knowledge' (2018) 47(6) *Research Policy* 1060–1069

Jeffrey L. Thorne, Hirohisa Kishino and Joseph Felsenstein, 'An Evolutionary Model for Maximum Likelihood Alignment of DNA Sequences' (1991) 33(2) *Journal of Molecular Evolution* 114–124

Jerry Thursby, Anne W. Fuller and Marie Thursby, 'US faculty patenting: Inside and outside the university' (2009) 38(1) *Research Policy* 14–25

Robert J. W Tijssen and Erik van Wijk, 'In Search of the European Paradox: an International Comparison of Europe's Scientific Performance and Knowledge Flows in Information and Communication Technologies Research' (1999) 28(5) *Research Policy* 519–543

Peter Tinnemann et al., 'Patenting of University and Non-University Public Research Organisations in Germany: Evidence from Patent Applications for Medical Research Results' (2010) 5(11) *PLOS ONE* e14059

Ruth Towse, 'The Quest for Evidence on the Economic Effects of Copyright Law' (2013) 37(5) *Cambridge Journal of Economics* 1187–1202

Maarten Truyens and Patrick Van Eecke, 'Legal Aspects of Text Mining' (2014) 30(2) *Computer Law & Security Review* 153–170

Hanns Ullrich, 'Expansionist Intellectual Property Protection and Reductionist Competition Rules: A TRIPS Perspective' (2004) 7(2) *Journal of International Economic Law* 401–430

Hanns Ullrich, 'Patent Protection in Europe: Integrating Europe into the Community or the Community into Europe?' (2002) 8(4) *European Law Journal* 433–491

P. Umesh et al., 'Programming Languages for Synthetic Biology' (2010) 4(4) *Systems and Synthetic Biology* 265–269

Garrett Upstill and Thomas H. Spurling, 'Adjusting to Changing Times: CSIRO Since the 1970s' (2007) 9(2) *Innovation* 113–124

Garrett Upstill and David Symington, 'Technology Transfer and the Creation of Companies: the CSIRO Experience' (2002) 32(3) *R&D Management* 233–239

Cathy Urquhart and Walter Fernández, 'Using Grounded Theory Method in Information Systems: the Researcher as Blank Slate and Other Myths' (2013) 28(3) *Journal of Information Technology* 224–236

Mikko Välimäki, 'Copyleft Licensing and EC Competition Law' (2006) 27(3) *European Competition Law Review* 130–136

Geertrui Van Overwalle, 'Policy Levers Tailoring Patent Law to Biotechnology: Comparing U.S. and European Approaches Bend or Break - The Patent System in Crisis' (2011) 1(2) *UC Irvine Law Review* 433–514

Drieke Vandamme et al., 'Systems Biology-Embedded Target Validation: Improving Efficacy in Drug Discovery' (2014) 6(1) *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 1–11

Vikrant Narayan Vasudeva, 'A Relook at Sui Generis Software Protection Through the Prism of Multi—Licensing' (2013) 16(1-2) *The Journal of World Intellectual Property* 87–103

Diane Vaughan, 'Theorizing Disaster Analogy, historical ethnography, and the Challenger accident' (2004) 5(3) *Ethnography* 315–347

J. Craig Venter et al., 'The Sequence of the Human Genome' (2001) 291(5507) *Science* 1304–1351

Greg R. Vetter, 'Are Prior User Rights Good for Software' (2014) 23(3) *Texas Intellectual Property Law Journal* 251–312

Saurabh Vishnubhakat and Arti Rai, 'When Biopharma Meets Software: Bioinformatics at the Patent Office' (2015) 29(1) *Harvard Journal of Law & Technology* 206

John P. Walsh, Charlene Cho and Wesley M. Cohen, 'View from the Bench: Patents and Material Transfers' (2005) 309(5743) *Science* 2002–2003

Glenn D. Walters, 'Predicting Subsequent Citations to Articles Published in Twelve Crime-Psychology Journals: Author Impact versus Journal Impact' (2006) 69(3) *Scientometrics* 499–510

Edward C. Walterscheid, ‘Early Evolution of the United States Patent Law: Antecedents (Part 2), The’ (1994) 76(11) *J. Pat. & Trademark Off. Soc’y* 849

Richard Watt, ‘Copyright and Contract Law: Economic Theory of Copyright Contracts The Relationship between Copyright and Contract Law’ (2010) 18(1) *Journal of Intellectual Property Law* 173–206

Mike Weed, ‘"Meta Interpretation": A Method for the Interpretive Synthesis of Qualitative Research’ (2005) 6(1) *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*

Joel West, ‘Commercializing Open Science: Deep Space Communications as the Lead Market for Shannon Theory, 1960–73’ (2008) 45(8) *Journal of Management Studies* 1506–1532

Joel West, ‘How Open is Open Enough?: Melding Proprietary and Open Source Platform Strategies’ (2003) 32(7) *Research Policy* 1259–1285

Sally Weston, ‘Software Interfaces - Stuck in the Middle: The Relationship Between the Law and Software Interfaces in Regulating and Encouraging Interoperability.’ (2012) 43(4) *IIC International Review of Intellectual Property and Competition Law* 427–450

James A. Whittaker and Jeffrey M. Voas, ‘50 years of software: key principles for quality’ (2002) 4(6) *IT Professional* 28–35

James C. Wilgenbusch and David Swofford, ‘Inferring Evolutionary Trees with PAUP*’ (2003) 00(1) *Current Protocols in Bioinformatics* 6.4.1–6.4.28

Yonghong Wu, Eric W. Welch and Wan-Ling Huang, ‘Commercialization of University Inventions: Individual and Institutional Factors Affecting Licensing of University Patents’ (2015) 36-37(February-March) *Technovation* 12–25

John Kojiro Yasuda and Christopher Ansell, ‘Regulatory capitalism and its discontents: Bilateral interdependence and the adaptability of regulatory styles’ (2015) 9(2) *Regulation & Governance* 178–192

Doogab Yi, ‘Cancer, Viruses, and Mass Migration: Paul Berg’s Venture into Eukaryotic Biology and the Advent of Recombinant DNA Research and Technology, 1967–1980’ (2008) 41(4) *Journal of the History of Biology* 589–636

Achim Zeileis, Christian Kleiber and Simon Jackman, 'Regression Models for Count Data in R' (2008) 27(8) *Journal of Statistical Software* 1–25

Chen Wei Zhu, 'A regime of droit moral detached from software copyright?—the undeath of the 'author' in free and open source software licensing' (2014) 22(4) *International Journal of Law and Information Technology* 367–392

Chen Wei Zhu, "Copyleft' Reconsidered Why Software Licensing Jurisprudence Needs Insights from Relational Contract Theory' (2013) 22(3) *Social & Legal Studies* 289–308

Nicole Ziegler, Oliver Gassmann and Sascha Friesike, 'Why do Firms Give Away their Patents for Free?' (2014) 37(2) *World Patent Information* 19–25

PhD Theses

Marcus M. Dapp, *The Effects of Software Patent Policy on the Motivation and Innovation of Free and Open Source Software Developers* (PhD Thesis, ETH Zurich, 2009)

Jason Michael Davis, *Reconsidering Antarctic Bioprospecting through Territorialities of Science, Property, and Governance* (PhD Thesis, The Ohio State University, 2011)

Eleanor Flening, *30 Years After the Bayh-Dole Act: Rethinking the Australian Research Commercialisation Experience* (PhD Thesis, 2012)

Jim David Flowers, *Explaining policy differences as a function of diverse governance institutions* (PhD Thesis, Georgia Tech, 2016)

Janis Dawn Geary, *How the DNA Barcode Commons are Governed: Understanding how a Heterogeneous Global Community Shares Genetic Resources for Non-Commercial Use* (PhD Thesis, University of Alberta, 2017)

James Howison, *Alone Together: A Socio-technical Theory of Motivation, Coordination and Collaboration Technologies in Organizing for Free and Open Source Software Development* (PhD Thesis, Syracuse University, 2009)

Annette McLeod, *Returns on Investment: Considerations on Publicly Funded ICT Research and Impact Assessment* (PhD thesis, University of Melbourne, 2016)

Gabriel J. Michael, *"To Promote the Progress"? Explaining the Global Diffusion of Intellectual Property Law* (PhD Thesis, The George Washington University, 2014)

Hazel V. J. Moir, *Do Patent Systems Improve Economic Well-Being? An Exploration of the Inventiveness of Business Method Patents* (PhD Thesis, The Australian National University, 2008)

Bruce Rasmussen, *Creating and capturing value in the biopharmaceutical sector* (PhD Thesis, Victoria University, 2008)

Nicholas Matthew Weber, *A Framework for Analyzing the Sustainability of Peer Produced Science Commons* (PhD Thesis, University of Illinois at Urbana-Champaign, 2015)

Conference Papers

Cristina Cifuentes and Anne Fitzgerald, 'Copyright in Shareware Software Distributed on the Internet—the Trumpet Winsock Case' (Paper presented at *Proceedings of the 19th International Conference on Software Engineering* , 1997)

Margaret Oakley Dayhoff and Robert S. Ledley, 'Comprotein: A Computer Program to Aid Primary Protein Structure Determination' (Paper presented at *Proceedings of the December 4-6, 1962, Fall Joint Computer Conference* , 1962)

Tom Dedeurwaerdere, 'Databases, Biological Information and Collective Action' (Paper presented at *Social Informatics: An Information Society for all? In Remembrance of Rob Kling* , 2006)

Melanie Dulong de Rosnay and Andrés Guadamuz, 'Open Access to Biodiversity Scientific Data: A Comparative Study' (Paper presented at *17th International Consortium on Applied Bioeconomy Research ICABR Conference on Innovation and the Policy for the Bioeconomy* , June 2013)

Robert English and Charles M. Schweik, 'Identifying Success and Tragedy of FLOSS Commons: A Preliminary Classification of Sourceforge.net Projects' (Paper presented at *First International Workshop on Emerging Trends in FLOSS Research and Development, 2007. FLOSS '07* , May 2007)

Imed Hammouda et al., 'Open Source Legality Patterns: Architectural Design Decisions Motivated by Legal Concerns' (Paper presented at *Proceedings of the 14th International Academic MindTrek Conference: Envisioning Future Media Environments* , 2010)

Christopher G. Harris, Robert Arens and Padmini Srinivasan, 'Comparison of IPC and USPC Classification Systems in Patent Prior Art Searches' (Paper presented at *Proceedings of the 3rd International Workshop on Patent Information Retrieval* , 2010)

James Howison and James D. Herbsleb, 'Scientific Software Production: Incentives and Collaboration' (Paper presented at *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* , 2011)

David L. Parnas, 'Designing Software for Ease of Extension and Contraction' (Paper presented at *Proceedings of the 3rd International Conference on Software Engineering* , 1978)

Jeremy Pitt and Ada Diaconescu, 'Structure and Governance of Communities for the Digital Society' (Paper presented at *2015 IEEE International Conference on Autonomic Computing* , July 2015)

David P. Randall, E. Ilana Diamant and Charlotte P. Lee, 'Creating Sustainable Cyberinfrastructures' (Paper presented at *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* , 2015)

Isabel Rose Reich and Victoria C. Stodden, 'Software Patents as a Barrier to Scientific Transparency: An Unexpected Consequence of Bayh-Dole' (Paper presented at *The Seventh Annual Conference on Empirical Legal Studies (CELS 2012)* , November 2012)

Charles M. Schweik et al., 'Brooks' Versus Linus' Law: An Empirical Test of Open Source Projects' (Paper presented at *Proceedings of the 2008 International Conference on Digital Government Research* , 2008)

Klaas-Jan Stol, Paul Ralph and Brian Fitzgerald, 'Grounded Theory in Software Engineering Research: A Critical Review and Guidelines' (Paper presented at *Proceedings of the 38th International Conference on Software Engineering* , 2016)

Sherif Yacoub, Hany Ammar and Ali Mili, 'Characterizing a Software Component' (Paper presented at *International Workshop on Component-Based Software Engineering* , 1999)

Technical Papers

Iain M. Cockburn, Rebecca Henderson and Scott Stern, 'The Impact of Artificial Intelligence on Innovation' (Working Paper No 24449, National Bureau of Economic Research, 2017) <http://www.nber.org/papers/w24449>

Jorge L. Contreras, 'Patents and Internet Standards' (Global Commission on Internet Governance Paper No 29, Centre for International Governance Innovation, 15th April 2018) <https://www.cigionline.org/publications/patents-and-internet-standards>

Bronwyn H. Hall and Dietmar Harhoff, 'Recent Research on the Economics of Patents' (Working Paper No 17773, National Bureau of Economic Research, July 2012)

Josh Lerner, 'The New New Financial Thing: The Sources of Innovation Before and After State Street' (Working Paper No 10223, National Bureau of Economic Research, January 2004) <http://www.nber.org/papers/w10223>

Benjamin Mitra-Kahn et al., 'The economic impact of innovation patents' (Economic Research Paper No 5, IP Australia, May 2015)

Victoria Stodden, 'The Scientific Method in Practice: Reproducibility in The Computational Sciences' (MIT Sloan School Working Paper No 4773-10, MIT Sloan School of Management, 2010) <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1550193>

Web Resources

Biotechnology Australia, *Australian Bioinformatics Network Final Report* (2009) <http://australianbioinformatics.net/storage/downloads/Australian%20Bioinformatics%20Network%20Project%20-%20Final%20Report%20-%2020080118.pdf>

James E. Bottomley, *Kernel developers' position on GPLv3 [LWN.net]* (22nd September 2006) <http://lwn.net/Articles/200422/>

European Commission, *Communication from The Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* <https://eur-lex.europea.eu/legal-content/EN/TXT/?uri=celex%3A52015DC0192>

Joshua Elkington, *BLAST Algorithm. Contribute to elkoutwest/BLAST development by creating an account on GitHub* (April 2016) <https://github.com/elkoutwest/BLAST>

Free Software Foundation, *How to use GNU licenses for your own software* (March 2018) <https://www.gnu.org/licenses/gpl-howto.en.html>

Ronggui Huang, *RQDA: R-based qualitative data analysis* (2018) <http://rqda.r-forge.r-project.org/>

Open Source Initiative, *The Open Source Definition (Annotated) (Version 1.9)* (22nd March 2007) <https://opensource.org/osd>

Scott K. Peterson, *GPL Enforcement Action in Hellwig v. VMware Dismissed, With an Appeal Expected* (11th August 2016) opensource.com <https://opensource.com/law/16/8/gpl-enforcement-action-hellwig-v-vmware>

Staff Reporter, *Preamitha Licenses NIPT Tech From Illumina Under Settlement* (19th September 2018) <https://www.genomeweb.com/molecular-diagnostics/premaitha-licenses-nipt-tech-illumina-under-settlement#.W-OM-3ozZcA>

Germán Rodrıguez, *Generalized Linear Models* (2017) <http://data.princeton.edu/wws509/r/overdispersion.html>

Richard M. Stallman, *Special Report: Apple's New Look and Feel* () GNU's Bulletin, vol. 1 no. 5 <https://www.gnu.org/bulletins/bull5.html#SEC9>

Richard M. Stallman, *What is Free Software?* (12th June 2018) <https://www.gnu.org/philosophy/free-sw.en.html>

R Core Team, *R: A Language and Environment for Statistical Computing.* (2013) R Foundation for Statistical Computing, Vienna, Austria <http://www.R-project.org/>